

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 26, 2003, 19:40:47 ; Search time 14833.9 Seconds  
(without alignments)  
17396.489 Million cell updates/sec

Title: US-09-724-126A-1

Perfect score: 6308

Sequence: 1 gccagaattcgccagagg.....aatttgatttggtgtttt 6308

Scoring table:

IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*

1: gb\_ba.\*

2: gb\_hgt.\*

3: gb\_in.\*

4: gb\_om.\*

5: gb\_ov.\*

6: gb\_pat.\*

7: gb\_ph.\*

8: gb\_pl.\*

9: gb\_pr.\*

10: gb\_ro.\*

11: gb\_sts.\*

12: gb\_sy.\*

13: gb\_un.\*

14: gb\_vt.\*

15: em\_ba.\*

16: em\_fun.\*

17: em\_hum.\*

18: em\_in.\*

19: em\_mu.\*

20: em\_om.\*

21: em\_or.\*

22: em\_ov.\*

23: em\_pat.\*

24: em\_ph.\*

25: em\_pl.\*

26: em\_ro.\*

27: em\_sts.\*

28: em\_un.\*

29: em\_vt.\*

30: em\_hgt\_hum.\*

31: em\_hgt\_inv.\*

32: em\_hgt\_other.\*

33: em\_hgt\_mus.\*

34: em\_hgt\_pln.\*

35: em\_hgt\_rod.\*

36: em\_hgt\_mam.\*

37: em\_hgt\_vrt.\*

38: em\_sy.\*

39: em\_hgt\_hum.\*

40: em\_hgt\_mus.\*

41: em\_hgt\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	5250	83.2	5250	9	AY061886	AY061886 Homo sapi
2	5085.6	80.6	5128	9	AF525401	AF525401 Homo sapi
3	4435.4	70.3	6395	6	AR030784	AR030784 Sequence
4	4435.4	70.3	6395	6	AR121463	AR121463 Sequence
5	4435.4	70.3	6395	10	AF061555	AF061555 Mus muscu
6	3055.8	48.4	3059	6	AX714232	AX714232 Sequence
7	3055.8	48.4	3059	6	AK056441	AK056441 Homo sapi
8	2546.8	40.4	2550	6	BD156870	BD156870 Primer fo
9	2546.8	40.4	2550	9	AK027803	AK027803 Homo sapi
10	1378.6	21.9	5268	9	AY061884	AY061884 Homo sapi
11	1233.4	19.6	5265	10	AY061885	AY061885 Mus muscu
12	996.2	15.8	1001	6	AR030785	AR030785 Sequence
13	996.2	15.8	1001	6	AR121464	AR121464 Sequence
14	994.2	15.8	999	9	AF061556	AF061556 Homo sapi
15	981	15.6	6158	9	AB002347	AB002347 Human mRN
16	772.8	12.3	3502	10	BC031403	BC031403 Mus muscu
17	784.2	12.1	818	9	HSX521	Z47040 Human parti
18	651.6	10.3	2958	9	AK026998	AK026998 Homo sapi
19	647	10.3	712	6	BD149613	BD149613 Primer fo
20	515.2	8.2	2475	10	BC026391	BC026391 Mus muscu
21	503.8	8.0	166518	9	AC090514	AC090514 Homo sapi
22	487.8	7.7	2512	9	AK026948	AK026948 Homo sapi
23	354	5.6	190727	9	AC068724	AC068724 Homo sapi
24	314.8	5.0	1368	10	BC025617	BC025617 Mus muscu
25	290.4	4.6	235978	2	AC094209	AC094209 Rattus no
26	287.6	4.6	181602	10	AL935168	AL935168 Mouse DNA
27	279.4	4.4	61896	2	AC090992	AC090992 Homo sapi
28	246.8	3.9	807	6	BD079625	BD079625 Cancer-as
29	224	3.6	5027	10	AK122254	AK122254 Mus muscu
30	216.8	3.4	6381	3	AY094815	AY094815 Drosophil
31	215.2	3.4	910	10	MM1UBR2	AF067372 Mus muscu
32	215.2	3.4	168200	2	AC120631	AC120631 Rattus no
33	215.2	3.4	198946	10	AL844548	AL844548 Mouse DNA
34	212.2	3.4	107304	2	AC016274	AC016274 Homo sapi
35	212.2	3.4	164468	2	AC021899	AC021899 Homo sapi
36	191.6	3.0	800	6	BD079626	BD079626 Cancer-as
37	180.8	2.9	164468	2	AC021899	AC021899 Homo sapi
38	160.6	2.5	3980	9	HS1UBR4	AF067383 Homo sapi
39	160.2	2.5	3411	9	AK090874	AK090874 Homo sapi
40	159	2.5	455	6	BD071823	BD071823 Secreted
41	157.6	2.5	594	11	G79301	G79301 S210P6045RG
42	150.2	2.4	107304	2	AC016274	AC016274 Homo sapi
43	132	2.1	28834	2	AC017982	AC017982 Drosophil
44	132	2.1	152545	3	AC010922	AC010922 Drosophil
45	132	2.1	179016	3	AC018489	AC018489 Drosophil

#### ALIGNMENTS

RESULT 1  
AY061886  
LOCUS  
DEFINITION Homo sapiens ubiquitin ligase E3 alpha-I mRNA, complete cds.  
ACCESSION AY061886  
VERSION AY061886.1 GI:27434483  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 5250)  
AUTHORS Han,H.Q. and Kwak,K.  
TITLE Full-length human ubiquitin ligase E3 alpha-I (E3 alpha)  
JOURNAL Unpublished

REFERENCE	2 (bases 1 to 5250)
AUTHORS	Han, H.-Q. and Kwak, K.
TITLE	Direct Submission
JOURNAL	Submitted (05-NOV-2001) Research, Angen Inc., One Angen Center Drive, Thousand Oaks, CA 91320, USA
FEATURES	Location/Qualifiers
source	1..5250
	/organism="Homo sapiens"
	/mol_type="mRNA"
	/db_xref="taxon:9606"
CDS	1..5250
	/codon_start=1
	/product="ubiquitin ligase E3 alpha-I"
	/protein_id="AA32103.1"
	/db_xref="GI:27434484"
	/translation="MADEPAGGPTGMEISAEELPOTPORLASWWDQKQVDFYATFALHLLA OLPVEIYFAEMDPDLKEQESVOMSTPTLEWLFQGEDPDICLEKLGSAFOALCGRV FKSGEITYSCRDCAIDNPTKCLVNDQFQDSVHKHNRHYKMTSTGGGFCGDDTAEWTKG PFCVNEPGRAGTITKENSCKPNEEVIVQARKTFPSVIVKIVEMTIEWEKEGRVAPPELO IRKRENYCYLVNDKSHSDYHVIYSLQALDCELAEOALHTTAIDKEGRRAVKAGAY AACQAEKEDIKSHSENVSQHPLRHEVYLHSEIMAHQKFLALRGLSWMMKIMSYSDSFQI FCQACLRPEPDSNPCLISRLMLDAKLYKGARKILHELIFSSFFMEMEYKKLFAMEF VKKYKOLQKEIISDDHRSTISFALSQVMTFPTLARHLIEQNVTSVITETLLEVL EYLDNRKNFCQYSODKLGRRVAVICDLKYLISKPTIWTESLRWQFLEGFSFLKI LTCWQNEEIRROGVGHIEVDPPWEAAIAIQMLKNILLMFQEWACADELLIVAIKRE CHAVMRCSTFISSSKTVQVSGHSELSYRVEDLSYHILPRLSTAGLHVRLSR LGVASLHEPVSFEDQVEVLYEPLRCLVVAQVVAEMRRNGLSLSIQVYIQDVR CREEMDKOITIMLQIGASMDPNKFLLLVQRYELAEAFNKTTISTKDDQLIKOYNTLI EEMQLVLYIVGRYVGVGNVTKEEVTWREITHLICIEPMHSAIAKNLIPENNET GLENLVKATFKPGVSGHVELKDESLKDFNMVYFYVSKTOHSAEHMOKRKO ENKDEALPPPPPEFCPAPSKVINLNCDIMMYILRTVFERAIDTSLNTEGMLONA FHIALGLEEKQLOKQAPPEEYDFYHKASRSGSAMNIQMLLEKLGIPOLEGQ DMITWLOPMDPTKYREKSLVATTSGESIKNDEITHDKEAKERKKAEARLHR QIKMAQSLQKNFIETHKLYDNTSEMPGKEDSIMEESTPAVSYSRIALGPKRGP SVTEKEVLTICLQEQEYQKIENNANVLACVOKTALTQHRGKPIELSGEALDPLFM DPLAYTYTSGQVHWYACVQKYPEAVOLSSQORTHVLDLPLESEYLCPLCKSLC NTVPIIPILOPKINSENDAAIAOLLTLARWIQTVLARISGYNIHAKGNPIPEFN QWGDSTLEHSLISFGVESSIKYSIKEMWILFATITIRIGLVKPPDPERDPVPMKL TQSCFTPTQSIENLGDGKPLFGALNQHQHKLALMQFAVQIRITCPOVLIKHL VRLSVLPINPKISDETPCLLSIDLPHVGVAVLAPPSLYWDDPVDLPQSSVSSYNHL YLFHLITMAHMLQILTTDPTGLPAQVEDSEEAHSASSPFAETISQVTSGISGCDLP LFWYWSLKNGITPLVRCAAFLPHVILGVTPEELHTNSAEGEYSALCSYLSLPTNLFL LFOEYVNPQLLQWCAADPALLNCLKOKNTVVPYPRKNSLLELPDDYSCLNLQASH FPRPSRDDRERKPIVCLFCGAILCSNICQEIIVGEEVGCACIFHAIHCGAGVCVIFL IKIECRVLYVEGRKARGCAPPYLDEYGETDPLGKGNPLHLHSRERYRKLUHLWQWOHC IIEIRARSQETNQMLGFGFNWLLI"
BASE COUNT	1576 a 1026 c 1180 g 1468 t
ORIGIN	
Query Match	83.28; Score 5250; DB 9; Length 5250;
Best Local Similarity	100.0%; Pred. No. 0;
Matches 5250; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
Qy	696 ATGCGGACGAGGAGGCTGGAGGTACTTGAGAGGATGGAATTCAGCGCGGAGTTACCCCG 755
Db	1 ATGCGGACGAGGAGGCTGGAGGTACTTGAGAGGATGGAATTCAGCGCGGAGTTACCCCG 60
Qy	756 ACCCCTCAGCGTCTGGCATCTTGTGGGATCAGCAAGTTGATTTTTATACGTCTTCTTG 815
Db	61 ACCCCTCAGCGTCTGGCATCTTGTGGGATCAGCAAGTTGATTTTTATACGTCTTCTTG 120
Qy	816 CATCATTTGGCACAAATTTGGTCCAGAAATTTACTTTCTGCTGAATGGACCCAGACTTGGAA 875
Db	121 CATCATTTGGCACAAATTTGGTCCAGAAATTTACTTTCTGCTGAATGGACCCAGACTTGGAA 180
Qy	876 AAGCAGGAGGAAAGTGTACAAATGTCAAATTTCACTCCACTTGGAAATGGTACTTATTTTGGGA 935
Db	181 AAGCAGGAGGAAAGTGTACAAATGTCAAATTTCACTCCACTTGGAAATGGTACTTATTTTGGGA 240
Qy	936 GAAGATCCAGATATTTGCTTAGAGAAATTTGAAGCACAGTGGAGCATTTTCAGCTTTGTGGG 995
Db	241 GAAGATCCAGATATTTGCTTAGAGAAATTTGAAGCACAGTGGAGCATTTTCAGCTTTGTGGG 300





QY 4296 AAATCTCTGTGCAATCTGTGATCCCATATTTCTTTGCAACCTCAAAAGATAAAGAGT 4355  
DB 3601 AAATCTCTGTGCAATCTGTGATCCCATATTTCTTTGCAACCTCAAAAGATAAAGAGT 3660  
QY 4356 GAGAAATGAGATGCTCTTGTCTCAACTTTTGACCCCTGGCAGCGGTGGATACAGACTGTTCTG 4415  
DB 3661 GAGAAATGAGATGCTCTTGTCTCAACTTTTGACCCCTGGCAGCGGTGGATACAGACTGTTCTG 3720  
QY 4416 GCCAAGATATCAGGTTATATATATAGACATGCTAAAGGAGAAAACCAATTCCTATTTC 4475  
DB 3721 GCCAAGATATCAGGTTATATATATAGACATGCTAAAGGAGAAAACCAATTCCTATTTC 3780  
QY 4476 TTTAATCAAGGAATGGAGATTTCTACTTTGGAGTTCCATTCCTCCTGAGTTTGGCGTT 4535  
DB 3781 TTTAATCAAGGAATGGAGATTTCTACTTTGGAGTTCCATTCCTCCTGAGTTTGGCGTT 3840  
QY 4536 GAGTCTTTCGATTAATATTCAAATAGCATCAAGGAAATGGTATTCTCTTTGCCACAACA 4595  
DB 3841 GAGTCTTTCGATTAATATTCAAATAGCATCAAGGAAATGGTATTCTCTTTGCCACAACA 3900  
QY 4596 ATTTATAGAAATGGATTGAAAGTGCCACCTGATGAAAGGGATCCTCGAGTCCCATGCTG 4655  
DB 3901 ATTTATAGAAATGGATTGAAAGTGCCACCTGATGAAAGGGATCCTCGAGTCCCATGCTG 3960  
QY 4656 ACCTGGAGACCTCGGCTTTCACATATCCAGGCAATGGAATCTATTGGGAGATGAAGGA 4715  
DB 3961 ACCTGGAGACCTCGGCTTTCACATATCCAGGCAATGGAATCTATTGGGAGATGAAGGA 4020  
QY 4716 AAACCTCTGTTGGAGACACTTCAAAATAGGACAGCATATGTCGTGAAGCATTAATGCGAG 4775  
DB 4021 AAACCTCTGTTGGAGACACTTCAAAATAGGACAGCATATGTCGTGAAGCATTAATGCGAG 4080  
QY 4776 TTTGCAGTTGCACAGAGGATTACCTGTCTCAGTGTCTGTATACAGAAACATCTGTTCTG 4835  
DB 4081 TTTGCAGTTGCACAGAGGATTACCTGTCTCAGTGTCTGTATACAGAAACATCTGTTCTG 4140  
QY 4836 CTTCTATCAGTTGTTCTTCTTAACATAAAATCAGAAGATACACCATGCTTCTGCTATA 4895  
DB 4141 CTTCTATCAGTTGTTCTTCTTAACATAAAATCAGAAGATACACCATGCTTCTGCTATA 4200  
QY 4896 GATCTGTTTCATGTTTGGTGGGTGCTGTGTAGCATCCCATCTGTATTTGGATGAC 4955  
DB 4201 GATCTGTTTCATGTTTGGTGGGTGCTGTGTAGCATCCCATCTGTATTTGGATGAC 4260  
QY 4956 CCTGTGATCTGCAGCCTTCTTCAAGTGTCTTCTTCAACACCATTTATCTCTTCCAT 5015  
DB 4261 CCTGTGATCTGCAGCCTTCTTCAAGTGTCTTCTTCAACACCATTTATCTCTTCCAT 4320  
QY 5016 TTGATCACCATGGCACACATGCTTTCAGATACACTTACAGTAGACAGGCTTACCCCTT 5075  
DB 4321 TTGATCACCATGGCACACATGCTTTCAGATACACTTACAGTAGACAGGCTTACCCCTT 4380  
QY 5076 GCTCAGGTTCAAGAGACAGTAGAGGCTCATTCGCGCATCTTCTTTTTCGAGAAAT 5135  
DB 4381 GCTCAGGTTCAAGAGACAGTAGAGGCTCATTCGCGCATCTTCTTTTTCGAGAAAT 4440  
QY 5136 TCTCAATATACAGTGGCTCCATTTGGGTGATATTTCTGCTGTATTTGGGTCTCA 5195  
DB 4441 TCTCAATATACAGTGGCTCCATTTGGGTGATATTTCTGCTGTATTTGGGTCTCA 4500  
QY 5196 CTGAAGATGGCATCACCCCTTATCTTCTGCTGTGTATTTTTCACATATTACTT 5255  
DB 4501 CTGAAGATGGCATCACCCCTTATCTTCTGCTGTGTATTTTTCACATATTACTT 4560  
QY 5256 GGGTAACTCCGCTGAGGAACATGCATACCAATTTCTGAGAGGAGATACAGTGCATC 5315  
DB 4561 GGGTAACTCCGCTGAGGAACATGCATACCAATTTCTGAGAGGAGATACAGTGCATC 4620  
QY 5316 TGTAGCTATCTATCTTTACCTACAAATTTGTTCTGCTCTTCCAGGAATATTTGGATACT 5375  
DB 4621 TGTAGCTATCTATCTTTACCTACAAATTTGTTCTGCTCTTCCAGGAATATTTGGATACT 4680

QY 5376 GTAAGGCCCTTGCTCCAGAGGTGGTCTGCAGATCCTGCCTTACTAAACTGTTTGAAGCAA 5435  
DB 4681 GTAAGGCCCTTGCTCCAGAGGTGGTCTGCAGATCCTGCCTTACTAAACTGTTTGAAGCAA 4740  
QY 5436 AAAAACAACGTCGTGTCAGGTACCTAGAAAAGAAATAGTTTGTATAGAGCTTCTCTGATGAC 5495  
DB 4741 AAAAACAACGTCGTGTCAGGTACCTAGAAAAGAAATAGTTTGTATAGAGCTTCTCTGATGAC 4800  
QY 5496 TATAGCTGCCCTCCTGAATCAAGCTTCTCATTTTCAGGTGCCAGGGTCTGCAGATGATGAG 5555  
DB 4801 TATAGCTGCCCTCCTGAATCAAGCTTCTCATTTTCAGGTGCCAGGGTCTGCAGATGATGAG 4860  
QY 5556 CGAAAGCATCTGTCCTCTGCTGCTTTTCTGTGGGCTTATCTATGTCTCAGAACATTTGC 5615  
DB 4861 CGAAAGCATCTGTCCTCTGCTGCTTTTCTGTGGGCTTATCTATGTCTCAGAACATTTGC 4920  
QY 5616 TGCCAGGAAATGTGAACGGGAAGAGTTGGAGCTTGCATTTTTCACGCACCTCACTGT 5675  
DB 4921 TGCCAGGAAATGTGAACGGGAAGAGTTGGAGCTTGCATTTTTCACGCACCTCACTGT 4980  
QY 5676 GGAGCCGAGTCTGCATTTTCTTAAATCAGAGATGCCGAGTGGTCTGTTGAAGGT 5735  
DB 4981 GGAGCCGAGTCTGCATTTTCTTAAATCAGAGATGCCGAGTGGTCTGTTGAAGGT 5040  
QY 5736 AAAGCAGAGGCTGTCCTTATCCAGCTCCTTACTTGGATGAATATGGAGAACAGACCT 5795  
DB 5041 AAAGCAGAGGCTGTCCTTATCCAGCTCCTTACTTGGATGAATATGGAGAACAGACCT 5100  
QY 5796 GGCCTGAAGAGGGGCAACCCCTTCAATTTATCTGTCGAGGGTATCGGAGCTCCATTTG 5855  
DB 5101 GGCCTGAAGAGGGGCAACCCCTTCAATTTATCTGTCGAGGGTATCGGAGCTCCATTTG 5160  
QY 5856 GTCCTGGCAACACACTGCAATATAGAGAGATTGCTAGGAGCCAGACACTAATCAGATG 5915  
DB 5161 GTCCTGGCAACACACTGCAATATAGAGAGATTGCTAGGAGCCAGACACTAATCAGATG 5220  
QY 5916 TTATTTGGATTCAACTGGCAGTTACTGTGA 5945  
DB 5221 TTATTTGGATTCAACTGGCAGTTACTGTGA 5250

RESULT 2  
AF525401  
LOCUS Homo sapiens UBR1 E3a ligase mRNA, partial cds. PRI 01-JAN-2003  
DEFINITION  
ACCESSION AF525401  
VERSION AF525401.1 GI:27451603  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 5128)  
AUTHORS Dgany,O., Avidan,N., Delaunay,J., Krasnov,T., Shalmon,L.,  
Shalev,H., Eidelitz-Markus,T., Kapeluschnik,J., Cattani,D.,  
Pariente,A., Tulliez,M., Cretien,A., Schischmanoff,P.O.,  
Iolascon,A., Fibach,E., Koren,A., Rossler,J., Le Merer,M.,  
Yaniv,I., Zilov,R., Ben-Asher,E., Olender,T., Lancet,D.,  
Beckmann,J.S. and Tamary,H.  
TITLE Congenital dyserythropoietic anemia type I is caused by mutations  
in codanin-1  
JOURNAL Am. J. Hum. Genet. 71 (6), 1467-1474 (2002)  
MEDLINE 22340442  
PUBMED 12434312  
REFERENCE 2 (bases 1 to 5128)  
AUTHORS Ben-Asher,E.  
TITLE Direct Submission  
JOURNAL Submitted (27-JUN-2002) Molecular Genetics, The Weizmann Institute  
of Science and the Schneider Medical Center, Hertzl Street,  
Rehovot 76100, Israel  
FEATURES  
Location/Qualifiers  
source 1..5128  
/organism="Homo sapiens"

/mol_type="mRNA"		Db		361	TACAAGATGCATCTTCTACTGAGAGGGTTCTGTGACTGTGACACACAGAGGCATGG	420
/db_xref="taxon:9606"		Qy		1167	AAACCTGGCCCTTTTGTGTAAATCATGAACCTGGAAGACAGGTACTATATAAAGAAAT	1226
/chromosome="15"		Db		421	AAACCTGGCCCTTTTGTGTAAATCATGAACCTGGAAGACAGGTACTATATAAAGAAAT	480
/map="15q13"		Qy		1227	TCACGCTGTCGGTGAATGAAGAGGTAATTGTCCAGCCAGGAAATATTTTCCTTCAGTG	1286
/cell_type="erythroid"		Db		481	TCACGCTGTCGGTGAATGAAGAGGTAATTGTCCAGCCAGGAAATATTTTCCTTCAGTG	540
<1..>5128		Qy		1287	ATAAATATGTCGTAGAAATGACTATATGGAAGAGGAAAGAACTGCTCTCTGAACTC	1346
/codon_start=1		Db		541	ATAAATATGTCGTAGAAATGACTATATGGAAGAGGAAAGAACTGCTCTCTGAACTC	600
/product="UBR1 E3a ligase"		Qy		1347	CAGATAAGGAGAGAAATGAAGATACTATTGTGTCCTTTCAATGATGAACACCATTC	1406
/protein_id="AA014997.1"		Db		601	CAGATAAGGAGAGAAATGAAGATACTATTGTGTCCTTTCAATGATGAACACCATTC	660
/db_xref="GI:27451604"		Qy		1407	TATGACCACTCATATACAGCCCTACAAAGAGCTCTTGACTGTGAGTTCGAGAGGCCAG	1466
/translation="LPOTPORLASWDDQOVDFVTAFLHHLAQLVPEIYFAEMDPDLEK		Db		661	TATGACCACTCATATACAGCCCTACAAAGAGCTCTTGACTGTGAGTTCGAGAGGCCAG	720
QESVMSIFTPLEWLFEGDPDICLEKLGSAFOLCGRVFKASGETTISCRDCAIDP		Qy		1467	TTGCATACCACTGCCATTGACAAAGAGGTCGTGGGCTGTTTAAAGCGGAGCTTATGCT	1526
TCVLCMDCQFVSVHKHRYKMTSTGGFCDCGDETAWKTPFCVNEHPRGATIKEN		Db		721	TTGCATACCACTGCCATTGACAAAGAGGTCGTGGGCTGTTTAAAGCGGAGCTTATGCT	780
SCSPLEEVIVQDAKEKTPFVYKIVYEMTITWEEKELPPELOIREKNERYCYLVFNDEH		Qy		1527	GCTTGGCAGGAACCAAGGAATATAAGAGGTCATTTCAGAAATGCTCTCTCAACATCCA	1586
SOPHDIYSLQELCELAEALHTTAIDKEGRVAKGAYAACQAKEDIKSHSENV		Db		781	GCTTGGCAGGAACCAAGGAATATAAGAGGTCATTTCAGAAATGCTCTCTCAACATCCA	840
HSHLVEYLRCLVLAQVAEMWRNGLSLISQVYVQDVKCEEMWDKILMLQIGA		Qy		1587	CTTTCATGTAGAAGTATTACACTCAGAGATTTATGGCTCATCAGAAATTTGCTTTCGCTCT	1646
ISRLMLDAKLKYGARKILHELIFSSFFMMEYKKLFAMEFYKYKQLKEYISDDHD		Db		841	CTTTCATGTAGAAGTATTACACTCAGAGATTTATGGCTCATCAGAAATTTGCTTTCGCTCT	900
RSISITALSVMQFVTPFLARHLIEQNVISVITETLLEVLPEYLDNRNKNFQGYSDQ		Qy		1647	GGTTCCTGGATGAACAAAATTTATGAGCTATTTCAAAGTGACITTTAGGCAGATCTTTTGGCAA	1706
KLGRVAVTDLKYILISKPTITWTERMQFLGFRSFLIKLITCMQMEELRRQVGH		Db		901	GGTTCCTGGATGAACAAAATTTATGAGCTATTTCAAAGTGACITTTAGGCAGATCTTTTGGCAA	960
IEYDPDWEAAIATQMOLKNILLMQEWCADELLHVLAYKECHAVMYRGSTFISSSK		Qy		1707	GCATGCTTTAGAGAAAGAACCTGACTCGGAGAAATCCCTGCTCATAGCAGGTTAAATGCTT	1766
		Db		961	GCATGCTTTAGAGAAAGAACCTGACTCGGAGAAATCCCTGCTCATAGCAGGTTAAATGCTT	1020
TVVQSCGHSLETKSYRVSDLVLSIHLPKSLTAGLHVRLSRGAVSLRHEFYVDFDFO		Qy		1767	TGGGATGCAAAAGCTTTTATAAGGTGCCCGTAAGATCCTTCATCAATTTGATCTTCACGAGT	1826
VEVLVEYLRCLVLAQVAEMWRNGLSLISQVYVQDVKCEEMWDKILMLQIGA		Db		1021	TGGGATGCAAAAGCTTTTATAAGGTGCCCGTAAGATCCTTCATCAATTTGATCTTCACGAGT	1080
SLMDPNKFLILLVREYLAEAFNKLTISDKDLIKQYNLILBEMLVQLVLIYVGRVVP		Qy		1827	TTTTTTTATGGAGATGGAATACAAAAAATCTTTTGTCTATGGAATTTCTGAAGTATTATAA	1886
GYNVTKEEIVREILIHLCIEBPMHSAIAKLNLENNETGLENNVLIENVIATFKPGV		Db		1081	TTTTTTTATGGAGATGGAATACAAAAAATCTTTTGTCTATGGAATTTCTGAAGTATTATAA	1140
SGHGVYELNDESLKDFNMFYHYSKTHSKAEHMKRRKQENKDALPPPPPPEFCP		Qy		1887	CAACTGCAGAAAGAAATATATCAGTGATGATCATGACAGAGATATCTCTATAAGTGCACCTT	1946
AFSKVINLNCEDIMMYILRTVFERAIDITNSLMEQLQMAFHILALGLEBKQOLQK		Db		1141	CAACTGCAGAAAGAAATATATCAGTGATGATCATGACAGAGATATCTCTATAAGTGCACCTT	1200
APEEVTVPFYKASRLGSSAMNIOMLLEKLGIPQLEGKTMITWILQMEDTVKRLR		Qy		1947	TCAGTTTCAGATGTTTACTGTTCTACTCTGGCTCGACATCTTATTGAAGAGCAGATGTT	2006
EKSLIVATTSSESINKDEIITHDEKAERKKAEAARLHRQKIMQMSALQKNFET		Db		1201	TCAGTTTCAGATGTTTACTGTTCTACTCTGGCTCGACATCTTATTGAAGAGCAGATGTT	1260
HKLMDYNTSEMPGKEDSIEEESTPAVSDYSRIAUGPKRGSPVTEKEVLTCLIQEEQ		Qy		2007	ATCTCTGTCATTACTGAAACTCTGCTAGAAAGTTTACCCTGAGTACTGTGACAGAGCAAT	2066
EYKINENNVLSACVOKSTALTQHRGKTELSGEALDPLFMDPLAIGTYTSCGHVM		Db		1261	ATCTCTGTCATTACTGAAACTCTGCTAGAAAGTTTACCCTGAGTACTGTGACAGAGCAAT	1320
HAVCWQYFEAVQLSQQRHIVDLFQESGEYLCLPLCKSLCNTVPIITLPQPKINSE		Qy		2067	AAATTTCAACTTCCAGGGTTTATAGCCAGGACAAAATTTGGGAAGAGTATATGACAGTAATATGT	2126
NADALAOLTLAKWTQITVLARI SGYNIRHAKGENPIPIFFNOGMDSTLEFHSILSFG		Db		1321	AAATTTCAACTTCCAGGGTTTATAGCCAGGACAAAATTTGGGAAGAGTATATGACAGTAATATGT	1380
VSSIKYSNISKEMILFATTIYRIGLKVPPDERPRVPLFWSCAFTIQALENLLG		Qy		2127	GACCTTAAAGTATATCTGTATCAGCAAAACCCACAATATGGACAGAAAGATTAGAATTCGAC	2186
DEGKPLFGALQNRHNLKALMQFAVQRITCPVLIQKHLVRLIUSVLPNKSEDTP		Db		1381	GACCTTAAAGTATATCTGTATCAGCAAAACCCACAATATGGACAGAAAGATTAGAATTCGAC	1440
CLLSIDLFLVGLVDAFLAPSLYWDVDPDLPSSVSSYNHLYLFLHITWAHMLQILLLT		Qy		2187	TTCCCTTGAAGGTTTTCGATCTTTTGAAGATTCTTACCTGTATGAGGAAATGGAAGAA	2246
VDTGLPLAQGVDSSEAHSAFSFAEISQYTSIGSDIPGWLYLWSLKNIGTPIYLRC		Db		1441	TTCCCTTGAAGGTTTTCGATCTTTTGAAGATTCTTACCTGTATGAGGAAATGGAAGAA	1500
AALFHYLGLVTPPEELHTNSAEGYSALCSYLSLPNLFLLFOEYWDVVRPLRQWC		Qy				
ADPALLNCLKQNTVVYRPRKNSLIELEPDDYSCLLNOASHFRCPSADDERKHPVLC		Db				
LFCGAILCSQNICCOEIVNGEVEGACIFHALHCGAGVCIFLKI RECRVVLVEKGARGC		Qy				
ATPAPYLDIEGTEDPLKRGNPLHFLVSGTEAPLVHQHCL*		Db				
BASE COUNT 1540 a 1010 c 1137 g 1439 t 2 others		Qy				
ORIGIN		Db				
Query Match 80.6%; Score 5085.6; DB 9; Length 5128;		Qy		747	TTACCCAGACCCCTCAGCGTCTGGCATCTTGGTGGGATCAGCAAGTTGATTTTATCT	806
Best Local Similarity 99.8%; Pred. No. 0;		Db		1	TTACCCAGACCCCTCAGCGTCTGGCATCTTGGTGGGATCAGCAAGTTGATTTTATCT	60
Matches 5121; Conservative 1; Mismatches 6; Indels 3; Gaps 3;		Qy		807	GCTTTTCTGCATCATTTGGCACAATTTGGTGCAGAAATTTACTTTGCTGAAATGGACCCA	866
		Db		61	GCTTTTCTGCATCATTTGGCACAATTTGGTGCAGAAATTTACTTTGCTGAAATGGACCCA	120
		Qy		867	GACTTGGAAAACGAGAGGAAAGTGTACAAATGTCAATATTCACCTCCACTGGAATGGTAC	926
		Db		121	GACTTGGAAAACGAGAGGAAAGTGTACAAATGTCAATATTCACCTCCACTGGAATGGTAC	180
		Qy		927	TTATTGGAGAGATCCAGATATTTGCTTTAGAGAAATTTGAAGCAGTGGAGCATTTTCAG	986
		Db		181	TTATTGGAGAGATCCAGATATTTGCTTTAGAGAAATTTGAAGCAGTGGAGCATTTTCAG	240
		Qy		987	CTTTGTGGAGGGTTTCAAAAGTGAGAGACAAACCTATTCTTTCAGAGGATTGTGCAATT	1046
		Db		241	CTTTGTGGAGGGTTTCAAAAGTGAGAGACAAACCTATTCTTTCAGAGGATTGTGCAATT	300
		Qy		1047	GATCCAACATGTGTACTCTGTATGGACTGCTTCCAGGACAGTGTTCATATAAATCATCGT	1106
		Db		301	GATCCAACATGTGTACTCTGTATGGACTGCTTCCAGGACAGTGTTCATATAAATCATCGT	360
		Qy		1107	TACAAGATGCATCTTCTACTGGAGAGGGTTCTGTGACTGTGGACACACAGAGGCATGG	1166

QY	2247	ATCCGAAGACAGGTTGGGCAACACATTTGAAGTGGATCCTGATTTGGAGGCTGCCATTGCT	2306
DB	1501	ATCCGAAGACAGGTTGGGCAACACATTTGAAGTGGATCCTGATTTGGAGGCTGCCATTGCT	1560
QY	2307	ATACAGATGCAATTTGAAGAATATTTTACTCATGTTTCCAAGAGTGGTGGCTGTGATCAA	2366
DB	1561	ATACAGATGCAATTTGAAGAATATTTTACTCATGTTTCCAAGAGTGGTGGCTGTGATCAA	1620
QY	2367	GAACCTTTACTTTGGGCTTATAAGAATGTCAACAAAGCTGTGATGAGTGCAGTACCAGT	2426
DB	1621	GAACCTTTACTTTGGCTTATAAGAATGTCAACAAAGCTGTGATGAGTGCAGTACCAGT	1680
QY	2427	TTCATATCTAGTACAGACAGTAGTACAATCGTGTGGACATAGTTTGGAAACAAGTCC	2486
DB	1681	TTCATATCTAGTACAGACAGTAGTACAATCGTGTGGACATAGTTTGGAAACAAGTCC	1740
QY	2487	TACAGAGTATCTGAGGATCTTGTAAAGATATGATGATATGATGATGATGATGATGATGAT	2546
DB	1741	TACAGAGTATCTGAGGATCTTGTAAAGATATGATGATATGATGATGATGATGATGATGAT	1800
QY	2547	CTTCATGTACGTTTAAAGCAGGCTGGGTGCTGTTTCAAGACATGATGATGATGATGATGAT	2606
DB	1801	CTTCATGTACGTTTAAAGCAGGCTGGGTGCTGTTTCAAGACATGATGATGATGATGATGAT	1860
QY	2607	GAGGACTTTCAAGTAGAGGTACTAGTGGAAATATCCTTTTACGTTGCTGCTGCTGCTGCT	2666
DB	1861	GAGGACTTTCAAGTAGAGGTACTAGTGGAAATATCCTTTTACGTTGCTGCTGCTGCTGCT	1920
QY	2667	CAGGTTGCTGATGATGTCGGAAGAAATGATGATGATGATGATGATGATGATGATGATGAT	2726
DB	1921	CAGGTTGCTGATGATGTCGGAAGAAATGATGATGATGATGATGATGATGATGATGATGAT	1980
QY	2727	TACCAAGATGTTAAGTCAGAGAGAAATGATGATGATGATGATGATGATGATGATGATGAT	2786
DB	1981	TACCAAGATGTTAAGTCAGAGAGAAATGATGATGATGATGATGATGATGATGATGATGAT	2040
QY	2787	GGTGATCTTTAATGGATCCCAATTAAGTCTTGTGTTTACTGGTACTTCAGAGGTATGAAC	2846
DB	2041	GGTGATCTTTAATGGATCCCAATTAAGTCTTGTGTTTACTGGTACTTCAGAGGTATGAAC	2100
QY	2847	GCCGAGGCTTTTACAGACCATATCTACAAAGACACAGGATTTGATTAACAATATATAT	2906
DB	2101	GCCGAGGCTTTTACAGACCATATCTACAAAGACACAGGATTTGATTAACAATATATAT	2160
QY	2907	ACACTAATAGAGAAATGCTTCAGTCTCTCATATATTTGGTGGTGGTGGTGGTGGTGGT	2966
DB	2161	ACACTAATAGAGAAATGCTTCAGTCTCTCATATATTTGGTGGTGGTGGTGGTGGTGGT	2220
QY	2967	GGAGTGGGAAATGTGACCAAGAGAGGTCACATGAGAGAAATCATTCACCTTGTCTTTC	3026
DB	2221	GGAGTGGGAAATGTGACCAAGAGAGGTCACATGAGAGAAATCATTCACCTTGTCTTTC	2280
QY	3027	ATTGAACCCATGCCACACAGTGCATTTGCCAAAATTTTACCTGAGAAATGAATAATGAA	3086
DB	2281	ATTGAACCCATGCCACACAGTGCATTTGCCAAAATTTTACCTGAGAAATGAATAATGAA	2340
QY	3087	ACTGGCTTAGAGATGTCATAACACAAAGTGGCCACATTTAAGAACACAGGTCATCAGC	3146
DB	2341	ACTGGCTTAGAGATGTCATAACACAAAGTGGCCACATTTAAGAACACAGGTCATCAGC	2400
QY	3147	CATGGAGTTTATGAACATAAAGATGAATCACTGAAAGACTTCAATATCTACTTTTATCAT	3206
DB	2401	CATGGAGTTTATGAACATAAAGATGAATCACTGAAAGACTTCAATATCTACTTTTATCAT	2460
QY	3207	TACTCCAAAACCCAGCATAGCAAGGCTGAACATATGCAGAGAAAGAGAGAAACAAGAA	3266
DB	2461	TACTCCAAAACCCAGCATAGCAAGGCTGAACATATGCAGAGAAAGAGAGAAACAAGAA	2520
QY	3267	AACAAAGATGAAGATTTGCCGCCACACACCTCTCTGAAATTTCTGCTTCTGCTTCTG	3326
DB	2521	AACAAAGATGAAGATTTGCCGCCACACACCTCTCTGAAATTTCTGCTTCTGCTTCTG	2580

QY	3327	GTGATTAACTTCTCAACTGTGATATCATATGATGATGATGATGATGATGATGATGATGATG	3386
DB	2581	GTGATTAACTTCTCAACTGTGATATCATATGATGATGATGATGATGATGATGATGATGATG	2640
QY	3387	GCAATAGACACAGATTCTAACTTGTGACCGAGGATGCTCCAAATGGCTTTTTCATATT	3446
DB	2641	GCAATAGACACAGATTCTAACTTGTGACCGAGGATGCTCCAAATGGCTTTTTCATATT	2700
QY	3447	CTGGCATTTGGGTTTACTAGAAGAGAACACAGCTTCAAAAAGCTCCTGAAGAAGAGTA	3506
DB	2701	CTGGCATTTGGGTTTACTAGAAGAGAACACAGCTTCAAAAAGCTCCTGAAGAAGAGTA	2760
QY	3507	ACATTTGACTTTTATCATATAAGGCTTCAAGATTTGGGAGTTTCAAGCATGATACAAATG	3566
DB	2761	ACATTTGACTTTTATCATATAAGGCTTCAAGATTTGGGAGTTTCAAGCATGATACAAATG	2820
QY	3567	CTTTTGGAAAACTCAAAGGAATTTCCCAAGTTAGAAAGGCCAGAAAGACATGATAAGCTGG	3626
DB	2821	CTTTTGGAAAACTCAAAGGAATTTCCCAAGTTAGAAAGGCCAGAAAGACATGATAAGCTGG	2880
QY	3627	ATACTTCAGATGTTTGACACAGTGAAGCGATTAAAGAGAAAAATCTTTGTTTAAATTTAGCA	3686
DB	2881	ATACTTCAGATGTTTGACACAGTGAAGCGATTAAAGAGAAAAATCTTTGTTTAAATTTAGCA	2940
QY	3687	ACCACATCAGGATCGGAATCTATTAGAATGATGATGATGATGATGATGATGATGATGATG	3746
DB	2941	ACCACATCAGGATCGGAATCTATTAGAATGATGATGATGATGATGATGATGATGATGATG	3000
QY	3747	GAACGAAAAAGAAAAAGCTGAAGCTGCTAGCTACATCGCCAGAGATCATGGCTCAGATG	3806
DB	3001	GAACGAAAAAGAAAAAGCTGAAGCTGCTAGCTACATCGCCAGAGATCATGGCTCAGATG	3060
QY	3807	TCCTGCTTACAGAAAAAATCTTCAATGAAACTCATATAAACTCATGATGATGATGATGATG	3866
DB	3061	TCCTGCTTACAGAAAAAATCTTCAATGAAACTCATATAAACTCATGATGATGATGATGATG	3120
QY	3867	ATGCTTGGGAAAGAGATTTCCATTTATGAGAGAGAGAGACCCAGCATGATGATGATGATG	3926
DB	3121	ATGCTTGGGAAAGAGATTTCCATTTATGAGAGAGAGAGACCCAGCATGATGATGATGATG	3180
QY	3927	TCTAGAAATGCTTTGGGTCTTAAACGGGTCCATCTGTTTACTGAAAGAGAGGTGCTGACG	3986
DB	3181	TCTAGAAATGCTTTGGGTCTTAAACGGGTCCATCTGTTTACTGAAAGAGAGGTGCTGACG	3240
QY	3987	TGCATCTTTGCCAAGAAACAGAGAGGTGAAAAATAGAAAAATTAATGCGCATGTTATTATCG	4046
DB	3241	TGCATCTTTGCCAAGAAACAGAGAGGTGAAAAATAGAAAAATTAATGCGCATGTTATTATCG	3300
QY	4047	GCCTGTGTCAGAAATCTACTGCCCTTAACCCAGCACAGGAGGAAACCCATAGAACTTCA	4106
DB	3301	GCCTGTGTCAGAAATCTACTGCCCTTAACCCAGCACAGGAGGAAACCCATAGAACTTCA	3360
QY	4107	GGAGAAAGCCCTAGACCCACTTTTTCATGGATCCAGACTTGGCATATGGAACCTTATACAG	4166
DB	3361	GGAGAAAGCCCTAGACCCACTTTTTCATGGATCCAGACTTGGCATATGGAACCTTATACAG	3420
QY	4167	AGCTGTGTCATGTAATGACGAGTGTGCTGGCAGAGTATTTTGAAGCTGTACAGCTG	4226
DB	3421	AGCTGTGTCATGTAATGACGAGTGTGCTGGCAGAGTATTTTGAAGCTGTACAGCTG	3480
QY	4227	AGCTCTCAGCAGCGCATTTTCTGACCTTTTGGCTGAAAGTGGAGAAATATCTTTTC	4286
DB	3481	AGCTCTCAGCAGCGCATTTTCTGACCTTTTGGCTGAAAGTGGAGAAATATCTTTTC	3540
QY	4287	CCTCTTTGCAAAATCTCTGTGCAATATCTGATCCCATTTATCTCTTTGCAACCTCAAAAG	4346
DB	3541	CCTCTTTGCAAAATCTCTGTGCAATATCTGATCCCATTTATCTCTTTGCAACCTCAAAAG	3600
QY	4347	ATAAAGATGAGAAATGCGATGCTCTGCTCAACTTTTGGCTTGGACCTGGCAGGTCGATAC	4406
DB	3601	ATAAAGATGAGAAATGCGATGCTCTGCTCAACTTTTGGCTTGGACCTGGCAGGTCGATAC	3660
QY	4407	ACTGTTCTGGCCAGAAATATCAGGTTTATATAAGACATGCTTAAAGAGGAGAAAAACCCAA	4466



308	GTGTACAGATGTC	CAATAC	TCTCCT	TTGGAGTGGT	ACTTATTTGGAGAGGAT	TCCGGATA	367
949	TTTTCCTTAGAGAA	TTGAACACAG	CGGAGCAT	TTTCAGCTTTTGGGAGGGTTT	CAAAA	1008	
368	TTTTCCTTAGAGAA	TTAAACACAG	TGGAGCGTTCCAGTTGCTGGGAAGGTTT	CANAA	427		
1009	GTGGAGAGACAAC	CTATCTTTG	CAGGAGTTGTC	CAATATGATCAACATG	GTGTTACTCTGTA	1068	
428	GTGGAGAAACAAC	ATATCTCTG	TAGGAGTTGTC	CAATATGATCAACATG	GTGTTACTCTGTA	487	
1069	TGCACTGCTTCC	AGGACAGT	TTTCATAAAAAT	CATCGTTACAGATG	GCATCTTACTG	1128	
488	TGGACTGCTTCC	AAAGTAGT	TTTCATAAAAAC	CGTTACAGATG	GCATCTTACTG	547	
1129	GAGGAGGGTCT	GTGACTGTG	GAGACACAGAG	CGCATGGAAAC	TGGCCCTTTTGTGTAA	1188	
548	GAGGGGCTTCT	GTGACTGTG	GAGACACAGAG	CGCTGGAAAC	TGGCCCTTTTGTGTGG	607	
1189	ATCATGAACCT	GGAAGCAGG	TACTATAAAAAG	AATTCACGCTGTCCGTTG	GAATGAAG	1248	
608	ATCAGAGCCT	TGAAGAGCAG	TACTACAAAAAG	AGCTTACATTTGCCATTTGAATGAAG	667		
1249	AGGTAAATG	TCCAGCCAG	AAATATTTTCC	TTCAGTGATAAAAT	TATGTCGTAGAAATGA	1308	
668	AGGTGATG	TGCTCAAGCC	AGAGAAATTTCC	TTTGGTGTAAAAAT	CATTTGTAGAAATGA	727	
1309	CTATATGGGAAG	AGGAAAAAAG	CTGCTCTGAAC	TCCAGATAAAGG	AGAAAAATGAAA	1368	
728	CTATATGGGAAG	AGGAAAAAAG	CTGCTCTGAAC	TGCGAGATAAAGG	AGAAAAATGAAAC	787	
1369	GATACATAT	TGTCCTTTT	CAATGATGAAC	ACCATTCATATGAC	CGCTCATATACAGCC	1428	
788	GATACATAT	TGTCCTTTT	CAACGATGAG	CACCATTCGTATGAT	CATGTATACAGTC	847	
1429	TACAAAGAGCT	TTTGACTGT	GAGCTGCGAGAG	CGCCAGTTGCATAC	CTGCGCATGACA	1488	
848	TGCAGAGAGCT	TAGATTTG	CGAGCTTG	CGAGAGCAGAGCT	TGCACAGCTTGC	907	
1489	AAGAGGCTCG	CGGGCTTTT	AAAGCGGAG	CTTATGCTGCTGCCAGG	AGCAAGGAAG	1548	
908	AAGAGGCTCG	CGGGCTTTT	AAAGCGGAG	CTTATGCTGCTGCCAGG	AGCAAGGAAG	967	
1549	ATATAAGAGCT	ATATCAGAA	AAATGCTCTCAAC	ATCCATCTCATGT	AGAAAGTATTACAT	1608	
968	ATATAAGAGCT	ATATCAGAA	AAATGCTCTCAAC	ATCCATCTCATGT	AGAAAGTATTACAT	1027	
1509	CAGAGATAT	TGGCTATC	AGAAATTTGCTTTG	CGCTTTGGTCTTCC	TGGATGAACAAATTA	1668	
1028	CCGTGGTTAT	GGCTACAGAA	ATTCGCTTGGCGCT	TGGCTTGGCTTGGCT	CCTGGATGAACAAATTA	1087	
1669	TGAGCTATT	CAAGTGACT	TTTAGCAGAT	CTTTTGGCCAGG	ATGCTTGGGATGCAAGCTTTATAAAG	1788	
1088	TGAGCTATT	CAAGTGACT	TTTAGCAGAT	CTTTTGGCCAGG	ATGCTTGGGATGCAAGCTTTATAAAG	1207	
1729	ACTCGAGAGAT	CCCTGCTCAT	AGCAGGTTATG	CTTTGGGATGCAAGCTTTATAAAG	1788		
1148	GCCTGAAAT	CCCTGCTCAT	AGCAGGTTATG	CTTTGGGATGCAAGCTTTATAAAG	1207		
1789	GTGCCCCGTA	AGATCCTTT	CATGAATGAT	CTTACGAGCTTTTATG	GGAGATGGAATACA	1848	
1208	GTGCCCCGTA	AGATCCTTT	CATGAATGAT	CTTACGAGCTTTTATG	GGAGATGGAATACA	1267	
1849	AAAACTCTTT	GCTATGGA	ATTTGTGAAG	TATTAATAAC	AACTGCGAAAAAGATATATCA	1908	
1268	AAAACTCTTT	GCTATGGA	ATTTGTGAAG	TATTAATAAC	AACTGCGAAAAAGATATATCA	1327	
1909	GTGATGAT	GTACAGAG	TATCTCTATA	CTTAACCTTT	CAGTTCAGATGTTTACTGTTTC	1968	
1328	CGACAGCACC	AGAGAGAG	CAATCTCCATA	ACCCTGTC	CGGTCAGATGCTCACCGTCC	1387	
1969	CTACTCTGG	CTCGACAT	CTTTTATGA	AGCAGAGATGTTAT	CTCTGTCATTTACTGAAACTC	2028	
1388	CGACTTGG	CGGCGAT	CTTTTATGA	AGCAGAGATGTTAT	CTCTGTCATTTACTGAAACTC	1447	



QY 3109 ACAAAGTGGCCACATTTAAGAAACACAGGTGCTATCAGGCCATGAGGTTTATGAACATAAAG 3168  
DB 2528 ACAAGTGGCCACATTTAAGAAACACAGGTGCTGCGGCCATGAGGTTTATGAATTTGAAG 2587  
QY 3169 ATGAATCACTGAAGAGCTTCAATATGTACTTTTATCATTAATCTCCAAAACCCAGCATAGCA 3228  
DB 2588 ATGAATCACTGAAGAGCTTCAATATGTACTTTTACCATTATTTCTAAACACAGCATAGCA 2647  
QY 3229 AGSCTGAACATATGCAAGAAAGGAGAAACAAAGAAACAAAGATGAAGCAATGCGCG 3288  
DB 2648 AGSCTGAACATATGCAAGAAAGGAGAAACAAAGAAATAAAGATGAAGCAATGCGCG 2707  
QY 3289 CACCACCACTCTCGAATTTGCGCTGCTTTTCAGCAAGATGAATTAACCTTTCTCAACTGTG 3348  
DB 2708 CGCCACCTCTCCAGAGTTCTGCGCTGCTTTTCAGCAAGTAGTCAACCTGCTAGCTGTG 2767  
QY 3349 ATATCATGATGATACATCTCAGACCGTATTTGAGCGGGCAATAGACACAGATCTCAACT 3408  
DB 2768 ATGTTATGATATACATCTCAGACCATCTTTGAGCGGGCAGTGGACACGGAGTCTTAATC 2827  
QY 3409 TGTGGACCGAAGGATGCTCCAATGGCTTTTCATATTTCTGGCATTTGGGTTTACTAGAAG 3468  
DB 2828 TGTGGACAGAGGATGCTGCAGATGGCTTCCATATATTGGCACTGGGCTTGGTGAAG 2887  
QY 3469 AGAAGCAACAGCTTCAAAAAGCTCCTGAAGAAAGTAACATTTGACTTTTATCATAGG 3528  
DB 2888 AGAAGCAGCAGCTTCAAGAAAGCTCCTGAAGAGGAGTGGCTTTGACTTTTACCATAAAG 2947  
QY 3529 CTTCAAGATTGGGAAGTTAGCCAT-----GAATATACAATGCTTTTGGAAAAAC 3579  
DB 2948 CTTCAAGATTGGGAAGTTAGCCATGAATGCTCAGAAATATACAAATGCTTTGGAAGAC 3007  
QY 3580 TCAAGGAATTCGCCAGTTAGAAAGCCAGAGGACATGATAACGTGGATCTTCAGATGT 3639  
DB 3008 TCAAGGAATTCGCCAATTAGAAGCCAGAGGACATGATAACGTACTCCAGATGT 3067  
QY 3640 TTGACACAGTGAAGCGATTAAAGAGAAAAATCTTTTAAATTGTAGCAACACATCAGGAT 3699  
DB 3068 TTGACACAGTGAAGCGATTAAAGAGAAAAATCTTTTAAATTGTAGTGGCAACCATTCAGGAC 3127  
QY 3700 CGGAATCTATTAGAATGATGAGATTACTCATGATTAAGAAAAAGCAGAACGAAAGAA 3759  
DB 3128 TGSAGTGCATTAGAGTGAAGGAGATTACTCATGATTAAGAAAAAGGAGCAAGCAAGAA 3187  
QY 3760 AAGCTCAAGCTGCTAGSCTACATCGCCAGAGGATCATGGCTCAGATGTCTGCTTACAGA 3819  
DB 3188 AAGCTGAGCGCTAGGCTTACCGCCAGAGATCATGGCCCAAGATGTCTGCTTACAGA 3247  
QY 3820 AAAACTTCATTTGAAACTCAATACTCATGTATGACAAATACATCAGAAATGCCCTGGAAAG 3879  
DB 3248 AAAACTTCATTTGAAACCCACAACATCATGTATGATTAATAGTCAGAAATACAGGAAG 3307  
QY 3880 AAGATTCCATTATGGAGAGAGAGACACCCAGCAGTCACTAGTACTCTAGAAATTGCTT 3939  
DB 3308 AAGACTCCATTATGGAGAGAGAGAGACACCTCAGCAGTCACTAGGCGCTCTAGAAATTGCTC 3367  
QY 3940 TGGCTCTAAACGGGCTCCTCTGTTACTGAAAGAGGAGTCTGAGCTGCATCTTGGCC 3999  
DB 3368 TGGGCTCTAAACGGGCGCGCTGTTACCGAAAGAGGAGTCTGAGCTGCATCTCTGCTGCC 3427  
QY 4000 AAGAAGAACAGGAGTGAATAATAGAAATTAATGCCATGTTATTTATGCGGCTGTGTCAGA 4059  
DB 3428 AAGAAGAACAGAGGTAAACTAGAAATAATGCCATGTTATTTGTCAGCATGTGTCAGA 3487  
QY 4060 AATCTACTGCTTTAACCCAGCAGACGGGAAAAACCCATAGAACTCTCAGGAGAACCCCTAG 4119  
DB 3488 AATCCACCGCCCTTAACCCAGCAGACAGGGAAGCCTGTGGACCCTTAGGGGAACACTGG 3547  
QY 4120 ACCCACTTTTCATGGATTCAGACTTGGCATATGGAATTAATAGAGAGCTGTGGTCATG 4179  
DB 3548 ACCCTCTTTTCATGGATTCAGACTTGGCATATGGAATTAATAGAGAGCTGTGGTCATG 3607  
QY 4180 TAATGACGCAAGTGTCTGGCAGAACTATTTTAAAGCTGTACAGCTGAGCTCTCAGCAGC 4239

DB 3608 TAATGCAATGCAAGTGTGCTGGCAGAAAGTATTTTGAAGCTGTGAGCTGAGCTCGCAGCAGC 3667  
QY 4240 GCATTTCATGTTGACCTTTTGTACTTTGGAAAGGAGAAATATCTTTGCCCTCTTTGGCAAT 4299  
DB 3668 GCATTTCAGTAGACCTGTTTGACCTGGAGAGCGGCGAGTACCTATGCTCCGCTCTGCAAGT 3727  
QY 4300 CTCTGTGCAATGCTGTGATCCCATTTATTCCTTTGCAACCTCAAAAGATAAACAAGTGAGA 4359  
DB 3728 CTCTGTGCAACACATGTCATCCCATCATCCCTTTGCAGCGCAGAAAGATCAACAGTGAGA 3787  
QY 4360 ATGCAGATGCTCTGCTCAACTTTTGACCTGGCAGAGTGGATACAGACTGTTCTGGCCA 4419  
DB 3788 ATGGGAGGCTCTTCTCAACTTTTGACCTGGCAGGTTGGATACAGACTGCTCTTGGCA 3847  
QY 4420 GAAATACAGGTTATAATATAACATGCTAAAGAGAAAC--CAATTCCTATTTCTT 4476  
DB 3848 GAAATTCGGGTTATAATATAAGCATGCTAAAGAGAAAGCCCGCAGCAGTTCCTGCTTGT 3907  
QY 4477 TTAATCAAGGAATGGGAGATTCTACTTTGGAGTTCCATTTCCATCTGAGTTTGGGCTTG 4536  
DB 3908 TTAATCAAGGAATGGGAGATTCAACTTTTGTAGTTTCATTTCCATCTGAGTTTGGAGTTC 3967  
QY 4537 AGTCTTCGATTAAATATTTCAATACATCAAGAAATGTTATTCTCTTTGGCACAAACAA 4596  
DB 3968 AGTCTTCGGTGAATAATTTCAAAATAGTATCAAGAAATGCTATTCTCTTCGCCACAAACAA 4027  
QY 4597 TTTATAGAAATGGATTGAAGTGCCACCTGATGAAGGATCCTCGAGTCCCATCTCA 4656  
DB 4028 TTTACAGAAATTTGGGCTGAAGTGCCCTCTGATGAATAGACCAAGTGGCCCATGATGA 4087  
QY 4657 CCTGAGCAGCTGCGCTTTTCACTATCCAGGCAATTTGAAATCTATTGGGAGATGAAGAA 4716  
DB 4088 CCTGAGCAGCTGTGCGTTTCCATCCAGGCAATCGAAACCTGTTGGGAGATGAAGAA 4147  
QY 4717 AACCTCTGTTTGGAGCACTTCAAAATAGGACGATTAAGTCTGTAAGAGCAATTAAGCACT 4776  
DB 4148 AACCTCTATTTTGGAGCACTTCAAAATAGACGATAGCGGTCTGAAGGCGCTAATGCACT 4207  
QY 4777 TTGCAGTTGCACAGAGGATTACCTCTCCTCAGTCTCTGATACAGAAACATCTGTTCTGTC 4836  
DB 4208 TTGCAGTTGCACAGAGGCTACCTGCCCTCAGGCTCTGATACACAAACATCTGGCTCGGC 4267  
QY 4837 TTCTATCAGTTGTTCTTCTTAACATAAATCAGAAATACAGATACACATGCTCTCTCTATAG 4896  
DB 4268 TCCTGTCAATTATCTTCTTCTTCACTGCAATCAGAAATACACAGGCTCTCTGTCTGTG 4327  
QY 4897 ATCTGTTTCATGTTTGTGGTGGTGTGTTTGTAGCATTCCTCTGTTGATTTGGATGACC 4956  
DB 4328 ATCTCTCCATGTTCTGTGTCGCGCAGTCTTAGCGTTCCCATCTGTTATTTGGGATGACA 4387  
QY 4957 CTGTTGATCTCGAGCTTCTTTCAGTTAGTTCTTCTCTATTAACACCTTTATCTCTCCATT 5016  
DB 4388 CCGTGGATCTCGACCGCTCGCCACCTAGTTCTTCTATTAACACCTCTATCTCTTCCATC 4447  
QY 5017 TGATCACCATGGCAGACATGCTTTCAGATACTTACTTACAGTAGACA-----CAG 5064  
DB 4448 TGATCACCATGGCAGACATGCTTCAGATACTCTTCAACAGATACAGATCTGTCTCCAG 4507  
QY 5065 GCCTACCCCTTGTCTCAGTTTCAAGAAAGACAGTGAAGGCTCATTTCCGCACTCTCTTCT 5124  
DB 4508 GGCCGCGCTTGTGTAGGGTGAAGAGGATAGTGAAGAGGCTCGCTGTCATCTGCTTCTCT 4567  
QY 5125 TTGCAGAAATTTCTCAATATACAAGTGGCTCCATTGGGTGTGATATTCTGCTGGTGGTATT 5184  
DB 4568 TTGTGGAAGTCTCGCAGCACACAGACGCGCTCACTGGGTGGGCTGCTCCCGCTGTATC 4627  
QY 5185 TGTGGGTCTCACTGAAGAAATGCACTACCCCTTATCTTCTGCTGTGCTGCTGCTGCTTCC 5244  
DB 4628 TGTGGCTCTCCTCAGGAACGCACTACCCCTTACCTCGCTGTGCTGCACTGCTTTCC 4687  
QY 5245 ACTATTTACTTTGGGTAACCTCGCGCTGAGGAACTGCATACCAATTTCTGCAGAGGAGAT 5304

Db 4688 ACTATTACTTGGAGTAGTCCGCGCTGAAGAAGCTGTTTGGCAATTCGTCTGAAGGAGAAT 4747  
QY 5305 ACAGTGACACTCTGTAGCTATCTATCTTTACTACAAATTTGTTCTCTGCTTCTCCAGGAAT 5364  
Db 4748 TCAGTGACACTCTGTAGCTATCTATCTTTACTCCACAAATTTGTTCTCTGCTTCTCCAGGAAT 4807  
QY 5365 ATTGGGATACTGTAAGGCCCTTGCTCCAGAGGTGGTGCGAGATCCCTGCTTACTAACT 5424  
Db 4808 ATTGGGATACCATGAAGCCCTTACTACAGAGGTGGTGCGAGATCCCTGCTTACTCAAGT 4867  
QY 5425 GTTTGAAGCAAAAGAACCCGTGGTACAGTACCCCTAGAAAAGAAATAGTTTGTATAGAGC 5484  
Db 4868 CTTTGAAGCAGAAAAGTGTCTGTCTAGGTACCCCTAGAAAAGAAATAGTTTGTATAGAGC 4927  
QY 5485 TTCCTGATGACTATAGCTGCTCTGTAATCAAGCTTCTCAATTCAGGTGCCACGGCTG 5544  
Db 4928 TTCCTGAGGACTACAGCTGCTTCTTAATCAGGCTTCTCAATTCAGGTGCCACGGCTG 4987  
QY 5545 CAGATGATGAGCGAAGACATCTGCTCTGCTTCTGCTTCTGCGGCTATACATGTTCTC 5604  
Db 4988 CAGATGATGAGCGAAGACATCTGCTCTGCTTCTGCTTCTGCGGCTATACATGTTCTC 5047  
QY 5605 AGAACATTTGCTGCCAGCAAAATTTGAACGGGGAAGAGTTGGAGCTTGCATTTTTCACG 5664  
Db 5048 AGACATCTGTTGCCAAGAAATAGTGAATGGGAAGAGGTGGAGCGTGGCTTTTCATG 5107  
QY 5665 CACTTCACTGTGGAGCCGGAGTGTGCATTTTCTTAAATAACAGAGAATGCCAGTGGTCC 5724  
Db 5108 CGCTTCATTTGTGTGCTGGAGTGTGCATTTTCTTAAATAACCGAGAATGCAGGTGGTCC 5167  
QY 5725 TGGTTGAAGTAAAGCCAGAGGCTGTGCCTATCCAGCTCCTTACTTGGATGAATATGAG 5784  
Db 5168 TGGTGAAGGAAAAGCCAGAGGCTGTGCCTATCCAGCCCTTACTTGGATGAATATGAG 5227  
QY 5785 AAACAGACCCTGGCTCGAAGAGGGCAACCCCTTCATTTATCTCGTGAGCGGTATCGGA 5844  
Db 5228 AAACAGACCAGGCTAAAGAGAGAAACCCACTTCATTTATCTCGGGAGCGGTATCGGA 5287  
QY 5845 AGCTCCATTTGGTCTGGCAACACACATGCATTATAGAGAGATGCTAGGAGCCCAAGAGA 5904  
Db 5288 AGCTGCATTTGGTCTGGCAACAGACATCTCATTTATAGAGAGATGCTCGGAGCCAGGAGA 5347  
QY 5905 CTAATCAGATGTTATTTGGATTCAACTGGCAGTGTACTGTAGCTCCAACTCTGCCTCAAG 5964  
Db 5348 CTAATCAGATGCTATTTGGATTAACTGGCAGTGTACTGTAGCTTCAGTTCTGCCTCAAG 5407  
QY 5965 ACAATCAAAATGACGACAGTAGTAAGGCTGATTCAAATATAGGAAACTTTCTGAGG 6024  
Db 5408 ACAATCAGTAGTACATCAATAAAGACTGTATCAAAATCTAGAGAATTTCTGAGG 5467  
QY 6025 GCTGGGAAAGTATTTGGAGGTCTTTTGTCTCCATGTCAGGTTTCACTTACATCAATAAAT 6084  
Db 5468 ACGGGGAAGTATTTGGAGGTCTTTTGTATCCATGTCAGATTCACACATTAATAAAT 5527  
QY 6085 ATTTCTTAAGGATGTTTGTCAATTAAGCAACATATGCTTCACAGGAAAAA-AGGAC 6143  
Db 5528 ATTCCTTAATGGAATATGCTTTCAATATCAACATTAAGCTTCAAGGGGAAAAAACAAGAC 5587  
QY 6144 ATAGATCAATCTGT- 6160  
Db 5588 ATAGATTAATGTTTATGTTCTAGAACACTAAAGAAATGCTTGTTCATCCAGGTCTCAT 5647  
QY 6161 ATGTGCTAGTATTTCCAGGAATTTATTTCCCTTTCATATTTTGTCTCATTTTCAATTTATTT 6220  
Db 5648 TTCTGCTAATATTTCCAGAAACTCTTTCCCTTCATACATGCTCTAGTTTCAATTCATAT 5707  
QY 6221 CATCCACTTGGTAGATGAAGTACGCTCAACAGTGTGTAGACATTTTATGTTGTTGTTAAC 6280  
Db 5708 CACCCACTGCTTAATGAGGTACATTAAGCATTTGTGGACATTTCTCCACTGCTGCTAAC 5767  
QY 6281 TCCTCTGCAATTTTGTATTTGGTGT 6307  
Db 5768 ATCTCTGCACCTTTGTATTTGGTGT 5794

RESULT 4  
AR121463  
LOCUS AR121463  
DEFINITION Sequence 1 from patent US 6159732.  
ACCESSION AR121463  
VERSION AR121463.1 GI:14105039  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 6395)  
AUTHORS Varshavsky,A. and Kwon,Y.Tae.  
TITLE Nucleic acid encoding mammalian Ubr1  
JOURNAL Patent: US 6159732-A 1 12-DEC-2000;  
FEATURES  
          source 1, 6395  
BASE COUNT 1802 a 1376 c 1519 g 1698 t  
ORIGIN  
  
Query Match 70.3%; Score 4435.4; DB 6; Length 6395;  
Best Local Similarity 86.3%; Pred. No. 0;  
Matches 4997; Conservative 0; Mismatches 721; Indels 69; Gaps 6;  
  
QY 590 TTCAGGGCCGTCGTAAAGTGTCTGCTCCTCTCTC-TCCGACGGGCCACAGGTTTCCGCT 648  
Db 8 TTCAGGGCCGTCGTAAAGTGTCTGCTCCTGTCGCGTCGCGCGCGGCCACAGGTTTCCGCT 67  
QY 649 TGCCTCTGCGCGGGGTGGCAACTGCAGCGCTCAGTTTCCCTCAAGATGGCGGACAGG 708  
Db 68 AGCTGCGCGCGGGGTGCGGAACCTGCGCGCTGTCGTTTCCCTTAAAGATGGCGGACGAG 127  
QY 709 AGGCTGGAGGTACTCAGAGGATGGAAATCAGCGCGAGTTACCCAGACCCCTCAGCGTC 768  
Db 128 AGATGGACGGCGCCGAGGAGGTGGAGTCAGCCCGGAGGCTCCCTTGGCGCCGACGGGC 187  
QY 769 TGGCATCTTGGTGGGATCAGCAAGTTGATTTTATATCTTCTTTCATCATTTTGGCAC 828  
Db 188 CGGCATCGTGGTGGGATCAGCAAGTTGATTTCTATACTGTTTCTTACATCATTTGGCAC 247  
QY 829 AATTGGTGGCAAGAAATTTACTTTGCTGAAATGGACCCAGACTTGGAAAAAGCAGAGGAAA 888  
Db 248 AATTAGTGCCAGAAATTTATTTTGTGAGATGGACCCAGATTGGAAAAAGCAAGAGAGA 307  
QY 889 GTGTCAAAATGCAATATTCCTCCACTGSAATGGTACTTATTTGGAGAGATCCAGATA 948  
Db 308 GTGTACAGATGCAATACTACTCTCTTGGAGTGGTACTTATTTGGAGAGGATCCGGATA 367  
QY 949 TTTGCTTAGAGAAATTTGAAGCACAGTGGAGCATTTTCAGCTTTTGGGAGGGTTTCAAAA 1008  
Db 368 TTTGCTTAGAGAAATTTAAACACAGTGGAGCGTTCCAGTTGTGTGGNAGGTTTCAAAA 427  
QY 1009 GTGGAGACACAACCTTATCTTGCAGGGATTGTGCAATTTGATCAACATGTGTACTGTGA 1068  
Db 428 GTGGAGAAACAACATATTCCTGTAGGATTGTGCAATTTGATCAACAGTGTGTCTGTGA 487  
QY 1069 TGGACTGCTTCCAGGACAGTGTTCATAAANAATCATCGTTACAGAGATGCATCTTCTACTG 1128  
Db 488 TGGACTGCTTCCAAAGTAGTGTTCATAAANAACCATCGTTACAAAGATGCATCTTCTACTG 547  
QY 1129 GAGGAGGGTCTCTGACTGTGGAGACACAGAGGCGATGGAAAACTGGCCCTTTTGTGTAA 1188  
Db 548 GAGGGGGTCTCTGACTGTGGAGACACAGAGCGTGGAAAACTGGCCCTTTTGTGTGG 607  
QY 1189 ATCATGAACCTGGAAGACAGGCTACTTATAAAAGAGAAATTCACCGTGTCCGTTGAATGAAG 1248  
Db 608 ATCAGAGGCTGGAAGACAGGCTACTTACAAAAGAGAGCTTACATTTGCCCATTTCAATGAAG 667  
QY 1249 AGGTAAATGTCGAAGCCAGCAAAATATTTCTTCACTGATATAAATATGCTGTAAGTAATGA 1308  
Db 668 AGGTGATTGCTCAAGCCAGGAGAAATATTCCTTCCGTTGATATAAATACATTTGTAGAATGA 727







QY	5665	CACCTTCACCTGCGAGCGGAGTCTGCATTTTCTCTAAATAATCAGAGATGCCGAGTGCTCC	5724
Db	5108	CGCTTCATTGTGGTCTGGAGTCTGCATTTTCTCTAAATAATCCAGAGATGCCAGGTGGTCC	5167
QY	5725	TGCTTGAAGGTAAGCCAGAGCGTGTGCCTATCCAGCTCCTTACTTGGATGAATATGAG	5784
Db	5168	TGGTGAAGMAAGCCAGAGCGTGTGCCTACCCAGCCCTTACTTGGATGAATATGAG	5227
QY	5785	AAACAGACCTGGCCTGAAGAGGGGCAACCCCTTCATTATCTCGTACGGTATCGGA	5844
Db	5228	AAACAGACCGGGCTTAAAGAGAGGAAACCCACTTCATTATCTCGGAGCGGTATCGGA	5287
QY	5845	AGCTCCATTGTCTGGCAACACACACTGCATTTATAGAAGATGCTAGGAGCAAGAGA	5904
Db	5288	AGCTGCATTGTGCTGGCAACACACTGCATTTATAGAAGATGCTCGGAGCAGGAGA	5347
QY	5905	CTAATCAGATGTTATTGTGATTCAACTGGCAGTTACTGTGAGCTCCAACCTCTCCCTCAAG	5964
Db	5348	CTAATCAGATGCTATTGTGATTCAACTGGCAGTTACTGTGAGCTTCAGTTCTCCCTCAAG	5407
QY	5965	ACAATCAAAATCACACAGTAGTAAGGCTGATTCAAAATTTATGGAATCTTCTGAGG	6024
Db	5408	ACAATCAGTGAATCAATAAAGACTGATCTAAAATTTCTAGAGAACTTCTGAGG	5467
QY	6025	GCTGGAAAGTATTGGAGGCTTTTGTCTCCATGTCCAGTTTCACTTACATCAATAAAAT	6084
Db	5468	ACGGGGAAGTATTGGAGGCTTTTGTATCCATGTCCAGATTTCACACACATTAATAAAAT	5527
QY	6085	ATTCTTCAATGAGTATCTCTTCAATTAGCAACATATGCTTCACAGGAAAAA-AGGAC	6143
Db	5528	ATTCTTCAATGAGTATCTCTTCAATTATCAACATAGCTTCAAGGGAAAAAACAAGAC	5587
QY	6144	ATAGATCAATCTGT-----TTTTTTT	6160
Db	5588	ATAGATTATGTTTATGTTCTAGACACATAAGAAATGCTGTTTATCAACAGTGCTAT	5647
QY	6161	ATGTGCTAGTATTTCCAGGAATTTATCCCTTCATAATTTGCTCATTTTATTTATTT	6220
Db	5648	TTCTGCTAATATTTCCAGAAAACTCTTCCCTTCATACTGCTTAGTTCAATTCATAT	5707
QY	6221	CATCCACTTGGTAGATGAAGTACGTCACGTCACAGATTGTAGACATTTATGTGTTGTTAAC	6280
Db	5708	CACCCACCTGGTAAATGAGGTACATTAAGCATTTCTGGACATTTCTCCATCTCGGTAAC	5767
QY	6281	TCCTTCGCAATTTTGTATTGGTGTTT	6307
Db	5768	ATCTCTGCACCTTTGTATTGGTGTTT	5794
RESULT 5			
AF061555		6395 bp mRNA linear ROD 10-JUL-1998	
LOCUS		Mus musculus ubiquitin-protein ligase E3-alpha (Ubr1) mRNA,	
DEFINITION		complete cds.	
ACCESSION		AF061555	
VERSION		AF061555.1 GI:3170886	
KEYWORDS			
SOURCE		Mus musculus (house mouse)	
ORGANISM			
		Mus musculus	
		Eukaryota;	
		Metazoa;	
		Chordata;	
		Craniata;	
		Vertebrata;	
		Euteleostomi;	
		Mammalia;	
		Eutheria;	
		Rodentia;	
		Sciurognathi;	
		Muridae;	
		Murinae;	
		Mus.	
REFERENCE			
AUTHORS		1 (bases 1 to 6395)	
		Kwon,Y.T., Reliss,Y., Fried,V.A., Hershko,A., Yoon,J.K., Gonda,D.K.,	
		Sangan,P., Copeland,N.G., Jenkins,N.A. and Varshavsky,A.	
TITLE		The mouse and human genes encoding the recognition component of the	
		N-end rule pathway	
JOURNAL		Proc. Natl. Acad. Sci. U.S.A. 95 (14), 7898-7903 (1998)	
MEDLINE		98318583	
PUBMED		9653112	
REFERENCE		2 (bases 1 to 6395)	
AUTHORS		Kwon,Y.T. and Varshavsky,A.	
TITLE		Direct Submission	

JOURNAL	Submitted (24-APR-1998) Division of Biology, 147-75, California		
	Institute of Technology, 1200 E. California Bl., Pasadena, CA		
	91125, USA		
FEATURES			
source	Location/Qualifiers		
	1. 6395		
	/organism="Mus musculus"		
	/mol_type="mRNA"		
	/db_xref="taxon:10090"		
	/chromosome="2"		
	/cell_line="MEL-C19"		
gene	1. 6395		
	/gene="Ubr1"		
5'UTR	1..114		
	/gene="Ubr1"		
CDS	115..5388		
	/gene="Ubr1"		
	/note="UBR1p; recognition component of the N-end rule		
	pathway"		
	/codon_start=1		
	/product="ubiquitin-protein ligase E3-alpha"		
	/protein_id="AAC40165.1"		
	/db_xref="GI:3170887"		
	/translation="MADREMGAERMDVSPPEPLAPORPASHWDDQVDFTYATFLHLHA		
	OLVPELYFAEMDPDLKQESVQMSILLTPLEWYLFGEQPDICLELKLKSGQTEAKWTG		
	FKSGEITISCRDCAIDPCVLCMDCFQSSVHKHRYKMTSTGGGFCDCGDTAEKWTG		
	PFCDVDEPGRAGTTKESLHCPLENEVIAQARRIFPSVIKYIVEMTWEEKEPELQ		
	IRKERNYCYVLFNDEHSHYVTSQALDCELAELAQHTTADKGGRAVAGVY		
	ATCQAEKEDIKSHSENQHPHLEVLVSHVMAHQFALRLGSMWNKIMSYSDPFOI		
	FCOACLEVEEGSENPCLISRLMDAKLYKGARKTLHELIFSSPFMEWYKLLFAMEF		
	VKYYKOLQKEYISDDHRSISITALSQVMTVPTLARHLIBEONVISVTEFLLEVL		
	EYLDNNKFNQGYSDQKLGRIYAVICDLKYLILSKPVIWTERLAQLEBGRFLKI		
	LTCMQMEETIRHQGYHIEVDPDWEAAIAIQMLKILLMFQWACDEDDLLLVAYKE		
	CHKAMSTGFMSTFQVLEVEYPLRCLVLAQVAEMWRNGLSLISQVFFYQDVK		
	LGALSRHEFPFDSFQVLEVEYPLRCLVLAQVAEMWRNGLSLISQVFFYQDVK		
	CREMYDKDIIMLOIGASIMDPNKLFLVLORELTDAFNKTIISKDODLIKQYNTLI		
	EMENLVLIYIVGRYVPGVGNVTRREVMREITHLLCEPHMPSAIARNLPENNET		
	GLENVINKVFKRPYGVGHVYELKSDNFMYFIYHSKTHSKAHMKRRKKQ		
	ENKDEALPPPPPEPFCAFQSVNLLSCDVMIIYILRTIFERAVQDTESNLWTEGLQMA		
	FHILALGLEEQQLQAPAEVEAFDFYHKASRLGSSAMNAQNIOMLLERLGIQLE		
	GOKDMITWLOQMTVRLREKSLVATTSGLECIKSEITHDEKEAERKKAAAR		
	LHRQKIMAOASALOKNFLETHKLMYDNTSEVTKGKEDSIWEEESTSAVSEASRIALGPK		
	RGPATKEVLTCLCQBEQEVKLENNAMVLISACVQKSTALTQHRGKPVDTLGETLDP		
	LFMDPDLAHGTYTGTSCGHMHAVCWQYFEAVLSQQRIHVDLDFEGLSEYCLPCK		
	SLCMTVAIIPLOQKINSEAAQLTLARWTQITQTLVRLISGNIHKAKGEAPVP		
	VLFNMGMDSTFEFHSILSGFQSVKYSNSTKEMVILFATTIYRIGLKVPPDLPDR		
	VPMMTWSTCAFTIOAENLDEGKPLFGALQNRHOSGLKALMQAVAOATCPOVLI		
	HKHLARLLSVILPNLQSENTGILLSVDLPHVLVGLAVLAFPSLYWDTVDLQSPLESS		
	YNHLFLHLITMAHMLQLLTTDDLSLSPPLAEGEDESEARCSAFEVESQITDG		
	LTCGAFQWLWLSLRNGITPYLRCAALLHYLLGAPPPEELFANSAGEFSALCYSL		
	SLPTNLFLLFOEYWDTRPLQRCMGDPALLSLKQKSAVVYPRKNSLIELPDSY		
	CLLNOASHFRCPRAADDERKHPVLCIFCGAILCSQNICQOEIVNGEEVACVFAHLC		
	GAGVCIFLKEICRECVLVGKARGCAYPAPVLDYGETDPGLKRGKNPLHLSRERYKL		
	HLVWQHQHIEIEIARSQETNOMLFGFNQMLL"		
3'UTR	5389..6395		
	/gene="Ubr1"		
polya_site	5429..5434		
	/gene="Ubr1"		
BASE COUNT	1802 a 1376 c 1519 g 1698 t		
ORIGIN			
	Query Match 70.3%; Score 4435.4; DB 10; Length 6395;		
	Best Local Similarity 86.3%; Pred. No. 0;		
	Matches 4997; Conservative 0; Mismatches 721; Indels 69; Gaps 6;		
QY	590	TTCAGGGCGCGTCTATAAAGTGTGCTCCTCTCTC-TCCGACCGGCCACAGGTTTCGGCT	648
Db	8	TTCAGGGCGCGTCTATAAAGTGTGCTCCTCTCTCCTCGCGCGGCCACAGGTTTCGGCT	67
QY	649	TGCCTCTGCGCGGGGTGCGCAACTGCAGGCGTTCAGTTTCCCTCAAGATGCGGCGAGG	708
Db	68	AGCTGCGCGCGGGGTGCGGAAGTTCGGGGGTCTGCTTTTCCCTTTAAGATGCGGCGAG	127
QY	709	AGGCTGGAGGTACTGAGAGGATGGAATCAGCGCGAGTTACCCCGACACCCCTCAGCGTC	768



Qy	2929	AGGCTCTCATATATCTGGTGAGCGTTATGTACCTGGAGTGGGAAATGTGACCAG	2988	Qy	4000	AAGAACAACAGGAGGTGAAAAATAGAAAAATATGCGATGGTATTATCGCCTGTGTCGAGA	4059
Db	2348	AGGCTCTCATATATCTGGTGAGAACGTTATGTACCTGGAGTGGGAAATGTGACCAG	2407	Db	3428	AAGAACAACAGGAGGTGAAAAATAGAAAAATATGCGATGGTATTATCGCCTGTGTCGAGA	3487
Qy	2989	AAGAGGTACAAATGAGAGAAATCAATTCATCTGCTTTGCAATGAACCCCATGCCACACAGTG	3048	Qy	4060	AATCTACTGCCTTTAAACCACACAGGGGAAAAACCCATAGAACTCTCAGGAGAAACCCCTAG	4119
Db	2408	AGGAGGTATATGAGAGAGATTACTCACTTACTTTGCAATGAGCCCATGCCACACAGTG	2467	Db	3488	AATCCACCGCCTTAACCCACACAGAGGGAAGCCTGTGGACCACTTAAAGGGAACACACTGG	3547
Qy	3049	CCATTGGCAAAATTTACTCGAGAAATGAAATATATGAACATGGCTTAGAGAAATGCATAA	3108	Qy	4120	ACCACTTTTCATGGATCCAGACTTGGCATATGGAACCTTATACAGGAAGCTGTGGTCATG	4179
Db	2468	CCATCGCAGAAACCTACTCGAGAAACCAAAATATGAACATGGCTTAGAGAAATGCATAA	2527	Db	3548	ACCTCTTTTCATGGATCCAGACTTGGCATATGGAACCTTATACAGGAAGCTGTGGTCATG	3607
Qy	3109	ACAAAGTGGCCACACATTTAAGAAACACAGGTGTATCAGGCCATGGAGTTTATGAACATAAAG	3168	Qy	4180	TAATGACACAGTGTCTGCGCAGAAATATTTTGAAGCTGTACAGCTGAGCTCTCAGCAGC	4239
Db	2528	ACAAAGTGGCCACACATTTAAGAAACACAGGTGTCTGCGGCCATGGAGTTTATGAATGAAG	2587	Db	3608	TAATGATGCAGTGTCTGCGCAGAAATATTTTGAAGCTGTGACGCTGAGCTCGCAGCAGC	3667
Qy	3169	ATGAATCACTGAAAGACTTCAATATGTACTTTTATCATTTACTTCCAAAACCCACAGCATAGCA	3228	Qy	4240	GCATTCATGTGTACCTTTTGTGACTTGGAAAGTGAGAAATATCTTTGCCCCTTTTGGCAAT	4299
Db	2588	ATGAATCACTGAAAGACTTCAATATGTACTTTTACCATTATCTTAAACACACAGCATAGCA	2647	Db	3668	GCATTCACGTAGACCTGTTTGGACTGGAGAGCGGAGTACCTATGCCCGCTCTGCAAGT	3727
Qy	3229	AGGCTGAACATATCAGAGAAAGAGAGAAACAAAGATGAAGCATTTGCCGC	3288	Qy	4300	CTCTGTGCAATACTGTGATCCCATTTATCTTTGCAACCTCAAAAGATAAACAGTGA	4359
Db	2648	AGGCTGAACATATCAGAGAAAGAGAGAAACAAAGATGAAGCATTTGCCGC	2707	Db	3728	CTCTGTGCAACACTGTATCCCATCATCTTTGCAAGCGCAGAAAGATCAACAGTGA	3787
Qy	3289	CACCAACACCTCTGAATTCGCCCTGCTTTCAGCAAAAGTGATTAACCTTCTCAACTGTG	3348	Qy	4360	ATGCAGATGCTTGTGCTCAACTTTTGACCTTGGCAGCGTGGATACACACTGTTCTGCCA	4419
Db	2708	CGCCACCTCTCCAGAGTTCTGCGCTGCTTTCAGCAAAAGTAGTCAACCTGCTCAGCTGTG	2767	Db	3788	ATGCGAGGCTCTTGTGCTCAACTTTTGACCTTGGCCCGGTGGATACACACTGCTCTGCCA	3847
Qy	3349	ATATCATGATGTACATTTCTCAGGACCGTATTTTGAGCGGCAATAGACACAGATTTCTAAT	3408	Qy	4420	GAATATCAGTATATATATAAGACATGCTAAAGGAGAAAAACC--CAATTCCTATTTTCT	4476
Db	2768	ATGTTATGATATATCTCTCAGGACCATCTTTTGAGCGGCGAGTGACACGGAGTTCTAATC	2827	Db	3848	GAATATCGGGTTATATATATAAGCATGCTAAAGGAGAAAGCCCGACAGTTCCTGCTTGT	3907
Qy	3409	TGTGGACCGAAGGATGCTCCAAATGGCTTTTCATATCTGGCATTTGGCTTTTACTAGAAG	3468	Qy	4477	TTAATCAAGGAATGGGAGATTTCTACTTTTGGAGTTTCCATTCCTCCTGAGTTTTGGCGTTG	4536
Db	2828	TGTGGACAGAGGATGCTCGAGATGGCTTCCATATATTGGCACTGGCTTGTCTGGAG	2887	Db	3908	TTAATCAAGGAATGGGAGATTTCAACTTTTGGAGTTTCCATTCCTCCTGAGTTTTGGAGTTC	3967
Qy	3469	AGAAGCAACAGCTTCAAAAAGCTCCTGAAGAAAGTAAACATTTTGACTTTTATCATAGG	3528	Qy	4537	AGTCTTCGATTAATATTCAAATAGCATCAAGGAAATGGTTATTTCTTTTGCCACACAA	4596
Db	2888	AGAAGCAACAGCTTCAAAAAGCTCCTGAAGAAAGTAAACATTTTGACTTTTATCATAGG	2947	Db	3968	AGTCTTCGATTAATATTCAAATAGCATCAAGGAAATGGTTATTTCTTTTGCCACAAAC	4027
Qy	3529	CTTCAAGATGGGAAGTTCAGCCAT-----GAATATACAAATGCTTTTGGAAAAAC	3579	Qy	4597	TTTATAGATTTGATTTGAAGTGGCACCCTGATGAAGGGATCCTCGAGTCCCCATGCTGA	4656
Db	2948	CTTCAAGATGGGAAGTTCAGCCATGAATGCTCGAAGTATACAAATGCTTTTGGAAAGAC	3007	Db	4028	TTTACAGAAATGGCCTGAAAGTGCCTCTGATGAACACTAGACCCACAGTAGTCCCATGATGA	4087
Qy	3580	TCAAGGAATCCCGAGTTAGAAGCCAGAGGACATGATTAACGTTGGATCTTCAGATGT	3639	Qy	4657	CCTGGAGCAGCTTCGCTTTTCACTATCCAGGCAATTTGAAATCTATTGGGAGATCAAGGAA	4716
Db	3008	TCAAGGAATCCCGAATTAGAGGCCAGAGGACATGATTAACATGGATTCACAGATGT	3067	Db	4088	CCTGGAGCAGTGTGGCTTCACCATCCAGGCAATCGAAACCTGTTGGGAGATGAAGAA	4147
Qy	3640	TTGACACAGTGAAGCGATTAAGAGAAAAATCTGTTTAATTTAGCAACCCACATCAGGAT	3699	Qy	4717	AACCTCTGTTTGGAGCAGCTTCAAAATAGGACGATTAATGTTGAAAGCATTAATGCAAT	4776
Db	3068	TTGACACAGTGAAGCGATTAAGAGAAAAATCTGTTTAATTTAGTGTGGCAACCATTCAGGAC	3127	Db	4148	AACCTCTAATTTGGAGCAGCTTCAAAATAGACAGCATAGCGCTCTGAAGCGCTAATGCAAT	4207
Qy	3700	CGGAATCTATTAAGATGATGAGATTACTATGATTAAGAGAAAAAGCAGAAAGAAAGAA	3759	Qy	4777	TTGACAGTGCACAGAGGATTTACCTGTCTCAGTCCCTGATACAGAAACATCTGTTGCTC	4836
Db	3128	TGGAGTGCATTAAGAGTGAAGAGATTACTATGATTAAGAGAAAAAGGCAAGCGAAGAA	3187	Db	4208	TTGACAGTGCACAGAGGAGTACCTGCCCTCAGTCCCTGATACACAAACATCTGGCTCGGC	4267
Qy	3760	AAGCTGAAGCTGTAGGCTACATCGCCAGAGATCATGCTCAGATGTCTGCTTACAGA	3819	Qy	4837	TTCTATCAGTGTGTTCTTCTTAACATAAAATACAGAAATACACCATGCTTCTGCTCTATAG	4896
Db	3188	AAGCTGAAGCGCTTAGGCTTCATCGCCAGAGATCATGGCCAGATGTCTGCTTACAGA	3247	Db	4268	TCTGTCACTTATCTTCTTCTTAACTGCAATCAGAAATACACAGGCTTCTGCTCTGG	4327
Qy	3820	AAAACTTCAATGAAACTCATAACTCATGTATGACAATACATCAGAAATGCTGCTGGAAG	3879	Qy	4897	ATCTGTTTCATGTTTGGTGGTGTCTGTAGCATTTCCCATCTTGTATTTGGATGACC	4956
Db	3248	AAAACTTCAATGAAACCCCAAACTCATGTATGATTAACGTGAGAAATGAGGAGG	3307	Db	4328	ATCTCTTCCATGTTTCTGGTGGCGCAGTCTTAGCGTTTCCCATCTCTGATTTGGATGACA	4387
Qy	3880	AAGATTCATTTAGGAGAGAGAGACCCAGCAGTCACTGACTTACTGATGATTTGCTT	3939	Qy	4957	CTGTTGATCTGACGCTTCTTCTAGTTAGTCTTCTTCTTATACCACTTTTATCTTCTCAT	5016
Db	3308	AAGACTCCATTTAGGAGAGAGAGACCTTCAGCAGTCACTGAGGCTTCTAGAAATGCTC	3367	Db	4388	CGTGGATCTGACGCGCTGCCACTTAGTTCTTTCATATAACCACTCTATCTCTTCCATC	4447
Qy	3940	TGGTCTCTAAACGGGCTCCATCTGTTTACTGAAAGGAGGCTGCTGACGTGCATCTTGGC	3999	Qy	5017	TGATCACCATGSCACACATGCTTCAGTACTACTTACAGTAGACA-----CAG	5064
Db	3368	TGGCCCTTAACGGGGCCCGCTGTTTACCGAAAAAGGAGGTGCTGACGTGCATCTCTGCC	3427	Db	4448	TGATCACCATGSCGACATGCTTTCAGTACTCTTCTTACACAGATACAGATCTGTCTCCAG	4507
				Qy	5065	GCCTACCCCTTGCTCAGGTTCAAGAGACAGTGAAGAGGCTCAATTCGCCATCTTCTTCT	5124

Db	4508	GGCGCGCGCTTGCTGAGGGTGAAGAGAGATAGTCAGAGGCGTCGCTGTCATCTGCTTTCT	4567
Qy	5125	TTGCAGAAAATTCTCAATATACAAGTGGCTCCATTTGGGTGTGATATTTCCCTGGCTGGTATT	5184
Db	4568	TTGTGGAAGTGTCGCAGCACACAGAGCGCTCACTGGTGGCGGTCTCCGGCTGGTACC	4627
Qy	5185	TGTGGGTCTCACTGAAGAATGGCATCACCCCTTATCTTCGCTGTGCTGCAATTTTTC	5244
Db	4628	TGTGGCTCTCCCTGAGGAACGGCATACCCCTTACCTCCGCTGTGCTGCACTGCTTTTC	4687
Qy	5245	ACTATTACTTTGGGTAACTCCGCTGAGGAAC TGATACCAATTTCTGCAGAAGGAGAT	5304
Db	4688	ACTATTACTTTGGAGTAGCTCCGCTGAAGAAC TGTTGCAATTTCTGCTGAAGGAGAAT	4747
Qy	5305	ACAGTGCACTCTGTAGCTATCTATTACTTACATAAAATTTGTTCCCTCTCCACGAAT	5364
Db	4748	TCAGTGCACTCTGTAGCTATCTATTACTTACCAACAATTTGTTCCGCTTTTCCAGGAAT	4807
Qy	5365	ATTGGGATACTGAAGGCCCTTGCTCCAGAGTGGTGTGCAGATCTCTGCCCTTACTAAACT	5424
Db	4808	ATTGGGATACCAATGAAGCCCTTACTACAGAGTGGTGTGGAGATCTCGCTTACTCAAGT	4867
Qy	5425	GTTTGAAGCAAAAACACACCGTCCAGGTACCCCTAGAAAAAGAAATAGTTTATAGAGC	5484
Db	4868	CTTTTGAAGCAGAAAAGTGTGTGGTCAAGTACCCCTAGAAAAAGAAATAGTTTATAGAGC	4927
Qy	5485	TTCTGTATGACTATAGCTGCCTCCTGAATCAAGCTTCTCATTTTCAGGTGCCACGGTCTG	5544
Db	4928	TTCTGTAGGACTACAGCTGTCTTCTTAATCAGGCTTCTCACTTTAGGTGCCACGGTCTG	4987
Qy	5545	CAGATGATGAGCAAAAGCATCTGTCTCTGSCCTTTTCTGTGGGGCTATACTATGTTCTC	5604
Db	4988	CAGATGATGAGCAAAAGCATCTGTCTCTGTCTTTCTGTGGGGCATCTCTGTGTTCTC	5047
Qy	5605	AGAACATTTGCTGCCAGGAAATTTGTGAACGGGGAAGAGTTGGAGCTTGCAATTTTTCAG	5664
Db	5048	AGAACATCTTGTCCAGAAATATAGTAATGGGAAGAGAGTTGGAGCTGTGGTTTTTCATG	5107
Qy	5665	CACTTCACTGTGAGCGGAGTCTGCATTTTCTTAAAAATCAGAGATCCGAGTGGTCC	5724
Db	5108	CGCTCAATTTGGTGTGGAGTCTGCATTTTCTTAAAAATCCAGAGATCCAGGGTGGTCC	5167
Qy	5725	TGTTGAAGTAAAGCCAGAGGCTGTGCCTATCCAGCTCCTTACTTTGGATGAATATGGAG	5784
Db	5168	TGTTGAAGAAAGCCAGAGGCTGTGCCTATCCAGCCCTTACTTTGGATGAATATGGAG	5227
Qy	5785	AAACAGACCCTGGCCTGAAGAGGGCAACCCCTTCAATTTATCTCGTGAAGCGGTATCGGA	5844
Db	5228	AAACAGACCAGGGCTAAAGAGAGGAACCCACTTCAATTTATCTCGGAGCGGTATCGGA	5287
Qy	5845	AGCTCCATTTGGTCTGGCAACACACTGCAATTTATAGNAGAGATGCTAGAGGCCAAGAGA	5904
Db	5288	AGCTGCATTTGGTCTGGCAACACACTGCAATTTATAGNAGAGATGCTCGGAGCCAGAGA	5347
Qy	5905	CTAATCAGATGTTATTGGAATCAACTGGCAGTTACTGTGAGCTCCAACCTCGCTCAAG	5964
Db	5348	CTAATCAGATGCTATTGGAATTTAACTGGCAGTATTCTCTGAGCTTCAGTTCTGCCTCAAG	5407
Qy	5965	ACAATCAAAATACGACAGTAGTAAGGCTGATTCAAAAATATGGAANAATTTCTGAGG	6024
Db	5408	ACAATCATGATGACATCAATAATAAGACTGATCTAAAAATTTCTAGAGAATTTCTGAGG	5467
Qy	6025	GCTGGGAAATATTGGAGGTCTTTTGCTCCATGTCCAGGTTCACITTCATCAATAAAT	6084
Db	5468	ACGGGGAAATATTGGAGGTCTTTTGATCCATGTCCAGATTCACACACATTAATAAAT	5527
Qy	6085	ATTCTTTAATGGAGTATTGCTTTTCAATTAGCAACAATATGCTTCACAGGAAAAA - AGGAC	6143
Db	5528	ATTCTTTAATGGAAATATTGCTTTTCAATTTCAACAATATGCTTCACAGGAAAAACAGAC	5587
Qy	6144	ATAGATCAATCTGT-----TTT	6160

Db	5588	ATAGATTAAATGTTTTATGTTCTTAGAACACATAAGAAAAATGCTTGTTCATCCAAAGTGTCTAT	5647
Qy	6161	ATGTCGTAGTATTTCCAGGAATTTATTCCTCTCATAAATTTGTCCTCATTTTATTTATTTT	6220
Db	5648	TTCTGCTAATATTTCCAGAAACTCTCTTCCCTTCATAACTGTCCTAGTTTCATTTTCATAT	5707
Qy	6221	CATCCACTTGGTAGATGAAGTCACGTCACAAACAGTTCTGACAGATTTTATGTTGGTTAAC	6280
Db	5708	CACCCACTGTTAATGACGGTCACATTAAGCATTTCTGGACATTTCTCCATCTGGCTAAC	5767
Qy	6281	TCITCTGCAATTTTGTATTTGGTGT	6307
Db	5768	ATCTGCACTTTGTATTTGGTGT	5794
RESULT 6			
AX714232			
LOCUS	AX714232	Sequence 916 from Patent EPI293569.	3059 bp DNA linear PAT 15-APR-2003
DEFINITION			
ACCESSION	AX714232		
VERSION	AX714232.1	GI:29889184	
KEYWORDS			
SOURCE		Homo sapiens (human)	
ORGANISM		Homo sapiens	
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
AUTHORS		1 Isogai,T., Sugiyama,T., Otsuki,T., Wakamatsu,A., Sato,H., Ishii,S., Yamamoto,J.I., Isono,Y., Hio,Y., Otsuka,K., Nagai,K., Irie,R., Tamechika,I., Seki,N., Yoshikawa,T., Otsuka,M., Nagahari,K. and Masuho,Y.	
TITLE		Full-length cDNAs	
JOURNAL		Patent: Ep 1293569-A 916 19-MAR-2003; Helix Research Institute (JP) ; Research Association for Biotechnology (JP)	
FEATURES		Location/Qualifiers	
source		1..3059	
		/organism="Homo sapiens"	
		/mol_type="genomic DNA"	
		/db_xref="taxon:9606"	
BASE COUNT	969 a	546 c	694 g 850 t
ORIGIN			
Query Match	48.4%	Score 3055.8;	DB 6; Length 3059;
Best Local Similarity	99.9%;	Pred. No. 0;	
Matches 3057;	Conservative 0;	Mismatches 2;	Indels 0; Gaps 0;
Qy	680	GTCAAGTTTCCTCAAGATGCGGACGAGGAGGCTGGAGGTACTGAGAGGATGGAATCAG	739
Db	1	GTCAAGTTTCCTCAAGATGCGGACGAGGAGGCTGGAGGTACTGAGGATGGAATCAG	60
Qy	740	CGGGAGTTACCCAGACCCCTCAGCGTCTGGCATCTTGGTGGGATCAGCAAGTTCATTT	799
Db	61	CGGGAGTTACCCAGACCCCTCAGCGTCTGGCATCTTGGTGGGATCAGCAAGTTCATTT	120
Qy	800	TTATCTGCTTCTTGTGATCATTTGGCACAAATTTGGTCCAGAAATTTACTTTGCTCAAA	859
Db	121	TTATCTGCTTCTTGTGATCATTTGGCACAAATTTGGTCCAGAAATTTACTTTGCTCAAA	180
Qy	860	GGACCCAGACTTGGAAAAGCAGGAGGAAAGTGTCACAAATGTCACATATTCACCTCCACTGGA	919
Db	181	GGACCCAGACTTGGAAAAGCAGGAGGAAAGTGTCACAAATGTCACATATTCACCTCCACTGGA	240
Qy	920	ATGCTACTTATTTGGAGAGATCCAGATATTTGCTTTAGAGAAATTCAGACACAGTGGAGC	979
Db	241	ATGCTACTTATTTGGAGAGATCCAGATATTTGCTTTAGAGAAATTCAGACACAGTGGAGC	300
Qy	980	ATTTACGTTTGTGGGGGGTTTTCAAAAGTGGAGAGACAAACCTATTTCTTGCAGGGATTG	1039
Db	301	ATTTACGTTTGTGGGGGGTTTTCAAAAGTGGAGAGACAAACCTATTTCTTGCAGGGATTG	360
Qy	1040	TGCAATTTGATCCAAACATGTCTGCTGTATGGACTGCTTCCAGGACAGTGTTCATAAAA	1099



Db 361 TGCAATTGATCCAACTGTGTAAGTCTGATGACTGCTTCCAGGACAGTGTTCATAAAA 420  
QY 1100 TCATCGTTACAGATGCATCTCTACTGAGGAGGTTCTGTGACTGTGAGACACAGA 1159  
Db 421 TCATCGTTACAAGATGCATCTCTACTGAGGAGGTTCTGTGACTGTGAGACACAGA 480  
QY 1160 GGCATGGAACCTGGCCCTTTTGTGTAAATCATGAACCTGGAGAGCAGTACTATAAA 1219  
Db 481 GGCATGGAACCTGGCCCTTTTGTGTAAATCATGAACCTGGAGAGCAGTACTATAAA 540  
QY 1220 AGAGAAATTCACGCTGTCGTTGAATGAAGAGTAAATGTCCAGCCAGGAAATATTTCC 1279  
Db 541 AGAGAAATTCACGCTGTCGTTGAATGAAGAGTAAATGTCCAGCCAGGAAATATTTCC 600  
QY 1280 TTCAGTGAATAAATATGCTGTAGAAATGACTATATGGAGAGGAAAAAGAACTGCCTCC 1339  
Db 601 TTCAGTGAATAAATATGCTGTAGAAATGACTATATGGAGAGGAAAAAGAACTGCCTCC 660  
QY 1340 TGAACCTCCAGATAAGGAGAAAAATGAAGATACTATTGTGCTCTTTTCAATGATGAACA 1399  
Db 661 TGAACCTCCAGATAAGGAGAAAAATGAAGATACTATTGTGCTCTTTTCAATGATGAACA 720  
QY 1400 CCATTCATATGACACGCTCATATACAGCCTTACAAAGAGCTCTTACTGTGAGCTCGCAGA 1459  
Db 721 CCATTCATATGACACGCTCATATACAGCCTTACAAAGAGCTCTTACTGTGAGCTCGCAGA 780  
QY 1460 GGGCCAGTTGCATACCACCTGCCATGTACAAAGAGGTCGTGGGGCTGTTAAAGCGGAGC 1519  
Db 781 GGGCCAGTTGCATACCACCTGCCATGTACAAAGAGGTCGTGGGGCTGTTAAAGCGGAGC 840  
QY 1520 TTATGCTGCTGCCAGGAAGCAAGAGATATAAGAGTCATTCAGAAAAATGCTCTCA 1579  
Db 841 TTATGCTGCTGCCAGGAAGCAAGAGATATAAGAGTCATTCAGAAAAATGCTCTCA 900  
QY 1580 ACATCCACTTCATGTAGAAGTATTACACTCAGAGATTATGGCTCATCAGAAAAATTTGCTTT 1639  
Db 901 ACATCCACTTCATGTAGAAGTATTACACTCAGAGATTATGGCTCATCAGAAAAATTTGCTTT 960  
QY 1640 GGGCTTTGGTTCCTGGATGAACAAAAATATGAGCTATTCAAGTGACTTTAGGAGATCTT 1699  
Db 961 GGGCTTTGGTTCCTGGATGAACAAAAATATGAGCTATTCAAGTGACTTTAGGAGATCTT 1020  
QY 1700 TTGCCAAGCATGCCTTAGAGAAACCTGACTCGGAGATCCCTGCTCATAGCAGGTT 1759  
Db 1021 TTGCCAAGCATGCCTTAGAGAAACCTGACTCGGAGATCCCTGCTCATAGCAGGTT 1080  
QY 1760 AATGCTTTGGGATGCAAGCTTTATAAGGTGCGCGTAAAGTCCCTCATCAATTTGATCTT 1819  
Db 1081 AATGCTTTGGGATGCAAGCTTTATAAGGTGCGCGTAAAGTCCCTCATCAATTTGATCTT 1140  
QY 1820 CAGCAGTTTTTTTATGGAGATGGAATACAAAAAACHCTTTTGTCTATGGAATTTGTGAAGTA 1879  
Db 1141 CAGCAGTTTTTTTATGGAGATGGAATACAAAAAACHCTTTTGTCTATGGAATTTGTGAAGTA 1200  
QY 1880 TTATAAACACTGCAGAAAGATATATCAGTGATGATCATGACAGAGTATCTCTATAAC 1939  
Db 1201 TTATAAACACTGCAGAAAGATATATCAGTGATGATCATGACAGAGTATCTCTATAAC 1260  
QY 1940 TGCACCTTCAGTTCAGATGTTTACTGTTCTTACTCTGGCTCGACATCTTATTGAAGACA 1999  
Db 1261 TGCACCTTCAGTTCAGATGTTTACTGTTCTTACTCTGGCTCGACATCTTATTGAAGACA 1320  
QY 2000 GAATGTTATCTCTGCTACTTACAAACTCTGCTAGAAGTTTTTACCTGAGTCTGGACAG 2059  
Db 1321 GAATGTTATCTCTGCTACTTACTGAAACTCTGCTAGAAGTTTTTACCTGAGTCTGGACAG 1380  
QY 2060 GAACATAAATTCACCTCCAGGGTTATAGCCAGGACAAATTTGGGAGAGTATATGTCAGT 2119  
Db 1381 GAACATAAATTCACCTCCAGGGTTATAGCCAGGACAAATTTGGGAGAGTATATGTCAGT 1440  
QY 2120 AATATGTGACCTAAAGTATATCTCTGATCAGCAAAACCCACAATATGACAGAGAAATTAAG 2179  
Db 1441 AATATGTGACCTAAAGTATATCTCTGATCAGCAAAACCCACAATATGACAGAGAAATTAAG 1500

QY 2180 AATCCAGTTCCTTGAAGGTTTTTCGATCTTTTTTGAAGATTTCTTACCTGTATGTCAGGAAT 2239  
Db 1501 AATCCAGTTCCTTGAAGGTTTTTCGATCTTTTTTGAAGATTTCTTACCTGTATGTCAGGAAT 1560  
QY 2240 GGAAGAAATCCGAAGACAGAGTTGGGCAACACATTTGAAGTGGATCTGATTTGGAGGCTGC 2299  
Db 1561 GGAAGAAATCCGAAGACAGAGTTGGGCAACACATTTGAAGTGGATCTGATTTGGAGGCTGC 1620  
QY 2300 CATTTGCTATACAGATGCAATTTGAAGAATATTTTACTCATGTTTCCAGAGTGGTGTGCTTG 2359  
Db 1621 CATTTGCTATACAGATGCAATTTGAAGAATATTTTACTCATGTTTCCAGAGTGGTGTGCTTG 1680  
QY 2360 TGATGAAGACCTCTTACTTGTGGCTTATAAAGAATGTCAAAAGCTGTGATGAGTGCAG 2419  
Db 1681 TGATGAAGACCTCTTACTTGTGGCTTATAAAGAATGTCAAAAGCTGTGATGAGTGCAG 1740  
QY 2420 TACCAGTTTCATATCTAGTAGCAAGACAGTAGTAGTACAATCGTGTGACATAGTTTGAAC 2479  
Db 1741 TACCAGTTTCATATCTAGTAGCAAGACAGTAGTAGTACAATCGTGTGACATAGTTTGAAC 1800  
QY 2480 AAAGTCTACAGAGTATCTGAGGATCTTTGAAGCATACATCTGCCACTCTCTAGGACCT 2539  
Db 1801 AAAGTCTACAGAGTATCTGAGGATCTTTGAAGCATACATCTGCCACTCTCTAGGACCT 1860  
QY 2540 TGCTGGTCTTCATGTACGTTTAAAGCAGGCTGGGTGCTGTTTCAAGACTGCATGATTTGT 2599  
Db 1861 TGCTGGTCTTCATGTACGTTTAAAGCAGGCTGGGTGCTGTTTCAAGACTGCATGATTTGT 1920  
QY 2600 GTCTTTTGAGGACTTTTCAAGTAGAGTACTAGTGAATATCTTTACGTTGCTGCTGTT 2659  
Db 1921 GTCTTTTGAGGACTTTTCAAGTAGAGTACTAGTGAATATCTTTACGTTGCTGCTGTT 1980  
QY 2660 GGTTCGCCAGGTTTGTCTGAGATGTGGCGAAGAAATGGACTGTCTCTTATTAGCCAGT 2719  
Db 1981 GGTTCGCCAGGTTTGTCTGAGATGTGGCGAAGAAATGGACTGTCTCTTATTAGCCAGT 2040  
QY 2720 GTTTTATTACCAAGATGTTAAGTGCAGAGAGAAATGTATGATAAGATATCATGCT 2779  
Db 2041 GTTTTATTACCAAGATGTTAAGTGCAGAGAGAAATGTATGATAAGATATCATGCT 2100  
QY 2780 TCAGATTTGGTGCATCTTTAATGGATCCCAATAAGTCTTTTACTGGTACTTTCAGAGT 2839  
Db 2101 TCAGATTTGGTGCATCTTTAATGGATCCCAATAAGTCTTTTACTGGTACTTTCAGAGT 2160  
QY 2840 TGAACCTTCCGAGGCTTTTAAAGACCATATCTACAAAAGACCAGGATTTGATTAACA 2899  
Db 2161 TGAACCTTCCGAGGCTTTTAAAGACCATATCTACAAAAGACCAGGATTTGATTAACA 2220  
QY 2900 ATATAATACATAATAGAAATAATGCTTTCAGTCTCTCATATATGTTGGGTGAGCGTTA 2959  
Db 2221 ATATAATACATAATAGAAATAATGCTTTCAGTCTCTCATATATGTTGGGTGAGCGTTA 2280  
QY 2960 TGTACCTTGGAGTGGAAATGTGACCAAGAGAGTCAACATGACAGAAATCATTCACCT 3019  
Db 2281 TGTACCTTGGAGTGGAAATGTGACCAAGAGAGTCAACATGACAGAAATCATTCACCT 2340  
QY 3020 GCTTTGCTTGAACCCCATGCCACAGTGCATGCCAAAATTTTACCTGAGAATGAAGA 3079  
Db 2341 GCTTTGCTTGAACCCCATGCCACAGTGCATGCCAAAATTTTACCTGAGAATGAAGA 2400  
QY 3080 TAATGAAACTGGCTTTAGAGAAATGTCAATAACAAAGTGGCCACATTTAAGAAACCCAGGT 3139  
Db 2401 TAATGAAACTGGCTTTAGAGAAATGTCAATAACAAAGTGGCCACATTTAAGAAACCCAGGT 2460  
QY 3140 ATCAGGCCATGGAGTTTATGAACATAAAGATGAATCACTAGAAAGACTTCAATATGACTT 3199  
Db 2461 ATCAGGCCATGGAGTTTATGAACATAAAGATGAATCACTAGAAAGACTTCAATATGACTT 2520  
QY 3200 TTATCATTTACTCCAAAACCCAGCATTAGCAAGCTGAACATATGACAGAGAAAGAGAA 3259  
Db 2521 TTATCATTTACTCCAAAACCCAGCATTAGCAAGCTGAACATATGACAGAGAAAGAGAA 2580





|||||  
841 TTATGCTGCTTGCAGGAAGCAAGAGATATTAAGAGTCATTCAGAAATGTCCTCA 900  
QY  
1580 ACATCCACTTCATGTAAGAGTATTACACTCAGAGATTATGGCTCATCAGAAATTTGCTTT 1639  
Db  
901 ACATCCACTTCATGTAAGAGTATTACACTCAGAGATTATGGCTCATCAGAAATTTGCTTT 960  
QY  
1640 GCGTCTTGCTTCTGATGACAAATATAGCTATTCAAGTGACTTTAGGCAGATCTT 1699  
Db  
961 GCGTCTTGCTTCTGATGACAAATATAGCTATTCAAGTGACTTTAGGCAGATCTT 1020  
QY  
1700 TTGCCAAGCATGCTTTAGAGAAACCTGACTCGGAGATCCCTGTCTCTCATAGCAGGTT 1759  
Db  
1021 TTGCCAAGCATGCTTTAGAGAAACCTGACTCGGAGATCCCTGTCTCTCATAGCAGGTT 1080  
QY  
1760 AATGCTTTGGGATGCCAAGCTTTATAAGSTGCCGTAAGATCCTTCATGAATTTGATCTT 1819  
Db  
1081 AATGCTTTGGGATGCCAAGCTTTATAAGSTGCCGTAAGATCCTTCATGAATTTGATCTT 1140  
QY  
1820 CAGCAGTTTTTTATGGAGATGGAATACAAAACCTTTTGGCTATGGAATTTGGAAGTA 1879  
Db  
1141 CAGCAGTTTTTTATGGAGATGGAATACAAAACCTTTTGGCTATGGAATTTGGAAGTA 1200  
QY  
1880 TTATAAACACTGCAGAAAGATATATACGTGATGATCATGACAGAGTATCTCTATAAC 1939  
Db  
1201 TTATTAAACACTGCAGAAAGATATATACGTGATGATCATGACAGAGTATCTCTATAAC 1260  
QY  
1940 TGCACCTTTTCAGTTCAGATGTTTACTGTTCTCTACTCTGGCTCGACATCTTATTGAAGAGCA 1999  
Db  
1261 TGCACCTTTTCAGTTCAGATGTTTACTGTTCTCTACTCTGGCTCGACATCTTATTGAAGAGCA 1320  
QY  
2000 GAATGTTATCTCTGTCATTACTGAAACTCTGCTAGAAAGTTTTACCTGAGTACTTGGACAG 2059  
Db  
1321 GAATGTTATCTCTGTCATTACTGAAACTCTGCTAGAAAGTTTTACCTGAGTACTTGGACAG 1380  
QY  
2060 GAACATAAATCAACTCCAGGTTATAGCCAGGACAAATTCGGAGAGTATATGCACT 2119  
Db  
1381 GAACATAAATCAACTCCAGGTTATAGCCAGGACAAATTCGGAGAGTATATGCACT 1440  
QY  
2120 AATATGTGACCTAAAGTATATCTGATCAGCAAAACCCACAAATATGCACAGAAAGATTAAAG 2179  
Db  
1441 AATATGTGACCTAAAGTATATCTGATCAGCAAAACCCACAAATATGCACAGAAAGATTAAAG 1500  
QY  
2180 AATGCAAGTTCCCTGTAAGGTTTTGATGCTTTTTTGAAGATCTTTACCTGTATGCAAGGAAT 2239  
Db  
1501 AATGCAAGTTCCCTGTAAGGTTTTGATGCTTTTTTGAAGATCTTTACCTGTATGCAAGGAAT 1560  
QY  
2240 GGAAGAAATCCGAAGACAGTTGGGCAACACATTTGAAGTGGATTCCTGATGCGAGGCTGC 2299  
Db  
1561 GGAAGAAATCCGAAGACAGTTGGGCAACACATTTGAAGTGGATTCCTGATGCGAGGCTGC 1620  
QY  
2300 CATTTGCTATACAGATCAATTTGAAGAAATATTTTACTCATGTTCCAAAGAGTGGTGTCTTG 2359  
Db  
1621 CATTTGCTATACAGATCAATTTGAAGAAATATTTTACTCATGTTCCAAAGAGTGGTGTCTTG 1680  
QY  
2360 TGATGAAGAACTCTTACTTTGGCTTTATAAAGAAATGTCACAAAGCTGTGATGAGGTCAG 2419  
Db  
1681 TGATGAAGAACTCTTACTTTGGCTTTATAAAGAAATGTCACAAAGCTGTGATGAGGTCAG 1740  
QY  
2420 TACCAGTTTCATATCTAGTAGCAAGACAGTAGTACAAATCGTGTGGACATAGTTTGGAAAC 2479  
Db  
1741 TACCAGTTTCATATCTAGTAGCAAGACAGTAGTACAAATCGTGTGGACATAGTTTGGAAAC 1800  
QY  
2480 AAAGTCCCTACAGATCTGAGGATCTTGTAGCATATACATCTGCCACTCTCTAGGACCTT 2539  
Db  
1801 AAAGTCCCTACAGATCTGAGGATCTTGTAGCATATACATCTGCCACTCTCTAGGACCTT 1860  
QY  
2540 TCGTGTCTTTCATGTTACGTTTAAAGCAGGCTGGTGTCTGTTTCAAGACTGCATCAATTTGT 2599  
Db  
1861 TCGTGTCTTTCATGTTACGTTTAAAGCAGGCTGGTGTCTGTTTCAAGACTGCATCAATTTGT 1920  
QY  
2600 GTCTTTTGGAGACTTTCAAGTAGAGTACTAGTGGAAATATCTTTACGTTGTCTGGTGT 2659  
Db  
|||||

1921 GTCTTTTCAGACCTTTCAAGTAGAGGTACTAGTGAATATCTTTACGTTGTCTGGTGT 1980  
QY  
2660 GGTTGCCCAGGTTGTCTGAGATGTGGCAGAAATGCAGCTGTCTCTTTATTAGCCAGGT 2719  
Db  
1981 GGTTGCCCAGGTTGTCTGAGATGTGGCAGAAATGCAGCTGTCTCTTTATTAGCCAGGT 2040  
QY  
2720 GTTTTATTACCAAGATGTTAAGTCAGAGAAATGATGATGAATATATCATCATGCT 2779  
Db  
2041 GTTTTATTACCAAGATGTTAAGTCAGAGAAATGATGATGAATATATCATCATGCT 2100  
QY  
2780 TCAGATTGGTGCATCTTTAATGGATCCCAATAAGTCTTTGTTACTTGGTACTTCAGAGGTA 2839  
Db  
2101 TCAGATTGGTGCATCTTTAATGGATCCCAATAAGTCTTTGTTACTTGGTACTTCAGAGGTA 2160  
QY  
2840 TGAACCTTCCCGAGGCTTTTAAACAAGACCATATCTACAAAAGACCAGGATTTGATTAAACA 2899  
Db  
2161 TGAACCTTCCCGAGGCTTTTAAACAAGACCATATCTACAAAAGACCAGGATTTGATTAAACA 2220  
QY  
2900 ATATAATACACTAATAGAAGAAATGCTTCAGGTCCTCATCTATATTTGGGTGAGGTTA 2959  
Db  
2221 ATATAATACACTAATAGAAGAAATGCTTCAGGTCCTCATCTATATTTGGGTGAGGTTA 2280  
QY  
2960 TGTACCTGGAGTGGAAATGTGACCAAAAGAGAGCTCAATGACAGAAATCATTCACAT 3019  
Db  
2281 TGTACCTGGAGTGGAAATGTGACCAAAAGAGAGCTCAATGACAGAAATCATTCACAT 2340  
QY  
3020 GCTTTGGATTTGAACCCCATGCCACAGTGCACATTTGCCAAAATTTACCTGAGAATGAAA 3079  
Db  
2341 GCTTTGGATTTGAACCCCATGCCACAGTGCACATTTGCCAAAATTTACCTGAGAATGAAA 2400  
QY  
3080 TAATGAACTGGCTTTAGAGAAATGTCATAAACAAGTGGCCACATTTAAGAAACCGAGTGT 3139  
Db  
2401 TAATGAACTGGCTTTAGAGAAATGTCATAAACAAGTGGCCACATTTAAGAAACCGAGTGT 2460  
QY  
3140 ATCAGGCCATGGAGTTTATGAACATAAAGATGAATCACTGGAAGACTTCAATATGACAT 3199  
Db  
2461 ATCAGGCCATGGAGTTTATGAACATAAAGATGAATCACTGGAAGACTTCAATATGACAT 2520  
QY  
3200 TTATCATTTACTCCAAAACCCAGCATAGCAAGCTCAACATATGCAGAGAAAAGAGAAA 3259  
Db  
2521 TTATCATTTACTCCAAAACCCAGCATAGCAAGCTCAACATATGCAGAGAAAAGAGAAA 2580  
QY  
3260 ACAAGAAAACAAGATGAAGCATTTGCCGCCACCACCTCTCTGAATTTCTGCCCTGCTTT 3319  
Db  
2581 ACAAGAAAACAAGATGAAGCATTTGCCGCCACCACCTCTCTGAATTTCTGCCCTGCTTT 2640  
QY  
3320 CAGCAAAAGTGAATTAACCTTCTCAACTGTGATATCATGATGATCATCTCAGAGACCTAT 3379  
Db  
2641 CAGCAAAAGTGAATTAACCTTCTCAACTGTGATATCATGATGATCATCTCAGAGACCTAT 2700  
QY  
3380 TGACGGGCAATAGACACAGATTTCTAAGTGTGGACCGAAGGATGCTTCCAAATGGCTTT 3439  
Db  
2701 TGACGGGCAATAGACACAGATTTCTAAGTGTGGACCGAAGGATGCTTCCAAATGGCTTT 2760  
QY  
3440 TCATATTTCTGGCATTGGGTTTTACTAGAGAGAAAGCAACAGCTTCAAAAAGCTTCCTGAAGA 3499  
Db  
2761 TCATATTTCTGGCATTGGGTTTTACTAGAGAGAGAAAGCAACAGCTTCAAAAAGCTTCCTGAAGA 2820  
QY  
3500 AGAAGTAACATTTGACTTTTATCATTAAGCTTCAAGTGTGGAACTTCAGCCATGAATAT 3559  
Db  
2821 AGAAGTAACATTTGACTTTTATCATTAAGCTTCAAGTGTGGAACTTCAGCCATGAATAT 2880  
QY  
3560 ACAAATGCTTTTGGAAAACCTCAAGGAATTTCCCCAGTTAGAGCCAGAGGACATGAT 3619  
Db  
2881 ACAAATGCTTTTGGAAAACCTCAAGGAATTTCCCCAGTTAGAGCCAGAGGACATGAT 2940  
QY  
3620 AACGTGATATCTCAGATGTTTGAACAGTGAAGCTTAAAGAGAAAATCTTGTGTTAAT 3679  
Db  
2941 AACGTGATATCTCAGATGTTTGAACAGTGAAGCTTAAAGAGAAAATCTTGTGTTAAT 3000  
QY  
3680 TGAGCAACACATCAGGATCGGATCTATTAAAGAAATGATGAGATTTACTCATGATAAAG 3738  
Db  
3001 TGAGCAACACATCAGGATCGGATCTATTAAAGAAATGATGAGATTTACTCATGATAAAG 3059



QY 2689 GAAGAAATGGACTGTCTCTATTAGCCAGCGTGTATTTATACCAAGATGTTAAAGTGCAGAG 2748  
Db 1501 GAGAAATGGACTGTCTCTATTAGCCAGCGTGTATTTATACCAAGATGTTAAAGTGCAGAG 1560  
QY 2749 AAGAAATGATGATAAAGATATCATCATGCTTCAGATTGTGTCATCTTTAATGATGCCCA 2808  
Db 1561 AAGAAATGATGATAAAGATATCATCATGCTTCAGATTGTGTCATCTTTAATGATGCCCA 1620  
QY 2809 ATAAGTCTTGTTACTGTTACTTTCAGAGGTATGAACCTTGCCGAGCGCTTTTAAACAGACCA 2868  
Db 1621 ATAAGTCTTGTTACTGTTACTTTCAGAGGTATGAACCTTGCCGAGCGCTTTTAAACAGACCA 1680  
QY 2869 TATCTACAAAGACCGAGATTTCATTAACAATAATATACATCAATTAAGAGAAATGCTTC 2928  
Db 1681 TATCTACAAAGACCGAGATTTCATTAACAATAATATACATCAATTAAGAGAAATGCTTC 1740  
QY 2929 AGTCTCTCATCTATATTGTGGGTGAGCGTTATGTACCTGAGTGGGAAATGTGACCAAG 2988  
Db 1741 AGTCTCTCATCTATATTGTGGGTGAGCGTTATGTACCTGAGTGGGAAATGTGACCAAG 1800  
QY 2989 AAGAGGTACAATGAGAGAAATCATTCACCTTGCTTTTGCAATGAACCCATGCCACACAGTG 3048  
Db 1801 AAGAGGTACAATGAGAGAAATCATTCACCTTGCTTTTGCAATGAACCCATGCCACACAGTG 1860  
QY 3049 CCATTGCCAAAATTTACCTGAGAAATGAATAATGAATGAATGCTTGAAGATGCTATAA 3108  
Db 1861 CCATTGCCAAAATTTACCTGAGAAATGAATAATGAATGAATGCTTGAAGATGCTATAA 1920  
QY 3109 ACAAACTGCCACATTTAAGAAACCCAGGTGTATCAGGCCATGGATTTATGAACATAAG 3168  
Db 1921 ACAAACTGCCACATTTAAGAAACCCAGGTGTATCAGGCCATGGATTTATGAACATAAG 1980  
QY 3169 ATGAATCACTGAAGAGCTTCAATATGTACTTTTATCATTTACTCCAAAACCCAGCATAGCA 3228  
Db 1981 ATGAATCACTGAAGAGCTTCAATATGTACTTTTATCATTTACTCCAAAACCCAGCATAGCA 2040  
QY 3229 AGGCTGAACATATGCAGAGAAAGAGGAGAAACAAAGAAATGAAGATGAAGATGCGCG 3288  
Db 2041 AGGCTGAACATATGCAGAGAAAGAGGAGAAACAAAGAAATGAAGATGAAGATGCGCG 2100  
QY 3289 CACCACCACTCTGTAATCTGCGCTGCTTTTCAGCAAGTGAATTAACCTTCTCAACTGTG 3348  
Db 2101 CACCACCACTCTGTAATCTGCGCTGCTTTTCAGCAAGTGAATTAACCTTCTCAACTGTG 2160  
QY 3349 ATATCATGATGATCTCTCAGGACCGTATTTGAGCGGGCAATAGACACAGATTTCTAACT 3408  
Db 2161 ATATCATGATGATCTCTCAGGACCGTATTTGAGCGGGCAATAGACACAGATTTCTAACT 2220  
QY 3409 TGTGGACCGAAGGATGCTCCAAATGGCTTTTCATATTTCTGGCATTTGGGTTTACTAGAAG 3468  
Db 2221 TGTGGACCGAAGGATGCTCCAAATGGCTTTTCATATTTCTGGCATTTGGGTTTACTAGAAG 2280  
QY 3469 AGAAGCAACAGCTTCAAAAAGCTCTCGAAGAAGTAACATTTGACTTTTATCATTAAGG 3528  
Db 2281 AGAAGCAACAGCTTCAAAAAGCTCTCGAAGAAGTAACATTTGACTTTTATCATTAAGG 2340  
QY 3529 CTTCAAGATTGGAAGTTCAGCCATGAATATCAAAATGCTTTTGGAAAACTCAAGGAA 3588  
Db 2341 CTTCAAGATTGGAAGTTCAGCCATGAATATCAAAATGCTTTTGGAAAACTCAAGGAA 2400  
QY 3589 TTCCCCAGTTAGAGGCCAAGGACATGATGAACGTTGATCTTCAGATGTTTGACACAG 3648  
Db 2401 TTCCCCAGTTAGAGGCCAAGGACATGATGAACGTTGATCTTCAGATGTTTGACACAG 2460  
QY 3649 TGAAGCGATTAGAGAAAAATCTGTTTAAATTTAGCAACACATCAGGATCGGAATCTA 3708  
Db 2461 TGAAGCGTTTAAAGAAAAATCTGTTTAAATTTAGCAACACATCAGGATCGGAATCTA 2520  
QY 3709 TTAAGAAATGATGATGATCTCATGATAAG 3738  
Db 2521 TTAAGAAATGATGATGATCTCATGATAAG 2550

RESULT 9  
AK027803  
LOCUS  
DEFINITION Homo sapiens cDNA FLJ14897 fis, clone PLACE1004743, weakly similar to PROBABLE N-END-RECOGNIZING PROTEIN.  
ACCESSION AK027803  
VERSION AK027803.1 GI:14042751  
KEYWORDS oligo capping; fis (full insert sequence).  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE 1  
AUTHORS Isogai,T., Ota,T., Hayashi,K., Sujiyama,T., Otsuki,T., Suzuki,Y., Nishikawa,T., Nagai,K., Sugano,S., Aotsuka,S., Yoshikawa,Y., Matsunawa,H., Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagahari,K., Masuho,Y. and Sasaki,N.  
TITLE NEDO human cDNA sequencing project  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 2550)  
AUTHORS Isogai,T. and Otsuki,T.  
JOURNAL Direct Submission  
COMMENT Submitted (10-MAY-2001) Takao Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986) NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology; cDNA library construction, 5'- & 3'-end one pass sequencing and clone selection: Helix Research Institute (supported by Japan Key Technology Center etc.) and Department of Virology, Institute of Medical Science, University of Tokyo.  
FEATURES  
Location/Qualifiers  
1..2550  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="PLACE1004743"  
/tissue\_type="placenta"  
/clone\_lib="PLACE1"  
/note="cloning vector: pME18SFL3"  
117..>2550  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="BAB55380.1"  
/db\_xref="GI:14042752"  
/translation="MTIWEKEKLPPELQIRKERNRYCYVLNDEHSHVDHYIYSIQR  
ALDCELAELAQHTTAIDKEGRVAVKAGYAAQCEAKEDIKSHENVSOHPLHVEVLHS  
EIMAHQKFAIRLGLSMNKMYSYSSDFRQIFCAQCLREEDPENPCILSRMLMDAKLY  
KGARKILHELIFSFFEMEMVKLFAMEFVKYKLOKEYISDDHDSRISITALSVMQ  
FTVPTLARHLIEONVISITETLLEVLPEYLDNRKNKFNQGYSDQKLGRRVAVICDL  
KYILISKPTIWTIRLRMQFLEGRFSFKILITCMQMEERIRVOGQHIQVDPWEAAIA  
IQMQLKNLILMFQWCACDELLIYAKCHKAVKRCSTFSISSRTVYVQSGHSLST  
KSYRVSDLVSIHLPLSRLTAGLHVRSLRHEFVSDFQVEVLVYPLRCL  
VLQAVVAEMRRNLSTLSIQVYQDVKCREEMDKDIIMQIGASLMDPNKFLILY  
LQRYELAEAKNTTSTQDQLIKOYNTLIEMLQVLIYTVGRYVPGVGNVTKVEVTM  
REIHLILCEPHSAIAKNLPENNETGLENVINKVATFKKPGVSGHGYELKDES  
LKDPMYFIHYSKTOHSAEHMQRKQKQNDKALPPPPPEFCAPSKVINLLNCD  
IMYILRTVFERADTSDNLTEGMLQAFHIALGLGLEEKQOLQAKPEEYTFDFYH  
KASRLGSSAMIQMLEKLGIPQLQEGOKDMITWILQMFDTVKRLREKSLIVATTSG  
SESTKNDIEITHDK"  
BASE COUNT 830 a 454 c 558 g 708 t  
ORIGIN  
Query Match 40.4%; Score 2546.8; DB 9; Length 2550;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 2548; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Qy 1189 ATCATGAACCTGGAGAGCAGGTACTATAAAGAGAAATTCACGCTGTCGCTGTAATGAAG 1248  
Db 1 ATCATGAACCTGGAGAGCAGGTACTATAAAGAGAAATTCACGCTGTCGCTGTAATGAAG 60  
Qy 1249 AGGTAATGTGCCAAGCCAGGAAATATTTTCCTTCAGTGATAAAATATGTCGTAGAATA 1308



```
QY 3469 AGAAGCAACAGCTTCAAAAGCTCCGGAAGAAGTAACATTTGACTTTTATCATAGG 3528
Db 2281 AGAAGCAACAGCTTCAAAAGCTCCGGAAGAAGTAACATTTGACTTTTATCATAGG 2340
QY 3529 CTTCAAGATTGGGAAGTTCCAGCCATGAATATACAAATGCTTTTGGAAAACTCAAGGAA 3588
Db 2341 CTTCAAGATTGGGAAGTTCCAGCCATGAATATACAAATGCTTTTGGAAAACTCAAGGAA 2400
QY 3589 TTCCCCAGTTAGAGGCCAGAGGACATGATACAGTGGATCTTCAGATGTTTGACACAG 3648
Db 2401 TTCCCCAGTTAGAGGCCAGAGGACATGATACAGTGGATCTTCAGATGTTTGACACAG 2460
QY 3649 TGAAGCGATTAGAGAAAAATCTTGTAAATGTAGCAACCATCAGGATCGGAATCTA 3708
Db 2461 TGAAGCGTTAGAGAAAAATCTTGTAAATGTAGCAACCATCAGGATCGGAATCTA 2520
QY 3709 TTAAGAAATGATGAGATTACTCATGATAAAG 3738
Db 2521 TTAAGAAATGATGAGATTACTCATGATAAAG 2550

RESULT 10
AY061884
LOCUS
DEFINITION Homo sapiens ubiquitin ligase E3 alpha-II mRNA, complete cds.
ACCESSION AY061884
VERSION AY061884.1 GI:27434479
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 5268)
AUTHORS Han, H.O. and Kwak, K.
TITLE Novel ubiquitin ligase E3 alpha-II
JOURNAL Unpublished
2 (bases 1 to 5268)
AUTHORS Han, H.O. and Kwak, K.
TITLE Direct Submission
JOURNAL Submitted (05-NOV-2001) Research, Amgen Inc., One Amgen Center
Drive, Thousand Oaks, CA 91320, USA
FEATURES
source
1..5268
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
1..5268
/codon_start=1
/product="ubiquitin ligase E3 alpha-II"
/protein_id="AAL32101.1"
/db_xref="GI:27434480"
/translation="MASELEPEVQALDRSLLECSAEIAGKWLQATDLTREVYQHLAH
YPKIYCRPNFPFQEDMLAHLVGLPMWELCGEDPAFGPKLEQANKPSHLGCRV
FKVGEPTSCRCADVPCLVMECLFGSIHRDHYRMTTSGGGFCDCGDTAEWREG
PYCQHELNTSIEEEDPLVLSIEDVIARTNIFAIRYAVAILTWKESLEPADL
EMVEKSDTYCMFNDEVHTYEQVIYTLQAVNCTOKEAIGFATTVDRGRSRYGD
FOYCEQAKSVIVRNTSROPKPLQVWHSIIYAHQNFGLKLLSWLSIIGYDGLRRI
LCQVGLQEPGENSLVDRMLSDLSKMGARSVTHQLFMSLLMDLKYKFLVYRF
AKNYQQLQGFEDDHERAVSVTALSQVFPTAPTLARMLITEENLSIIITKTFDAFL
HRDAQGRFERYTLQAFKFRVQSLILDLYVLSKPTSEWDELQKFLFEGDAFL
ELKCMQMDGHTYDGEQPIITLSICGHSVETRYCVSOEKYSIHLVPSRLLAGHLV
KKCLAVLMKCHGSDYDGEQPIITLSICGHSVETRYCVSOEKYSIHLVPSRLLAGHLV
LSKSEVAKFPELLPLSELSPMLIEHPLRCLVLCQAVHAGWRNRGFSLVNQIYYH
NVKCRREMDKVMQTLQGVSMWDPNHLMLISRFELYOIFSTPDYGRKFSSEITHK
DYVOQNTLIEMLYLIIMLVGRFSPGVQVNDATDEIKREIITHLSIKPMSESLVK
SUPEDENKTRGEMTIEVAHFPPKGLTGRVVELKPECAKFNLYFYHFSRAEQSKA
EPAORLKRQNRDIALPPPLPPCPPLFASLVNIIQSDVLMICMTILQWAVEHNGY
ANSENLQVRLHLIGMALQEBQHLNENVEEHWVFTFTQKISKPEAKPNSPSILAM
LETLQNAPLYLEHKDMIRWLKTFNAVKMRESSPTSPVAETEGTITMESSRDKRAE
RRKAEIRLAEIRLQMAQSEMRHFIDENKELFOOTLEDASTSAVLDSHPVSQMT
LTALGPAQTQVPEQRFVTCILCQEBQEVKVESRAMVLAAFVQRSYVLKNRSKFIQD
PEKYDPLFMPDLSCGTHTSSCGGHIMHAWCQRYFDSVQAEQRRQRLRLHTSYDVE
```

```
NGEFLCPLECLSNVTPILLPPRNIPIRNRLNFSNPDLTOWIRTIISOIKALOFRLK
EESPNNASTKNSENVDELQLPFEGFRPDRPKPIPYSESIKEMLTTFPGTATYKVLKH
FNEEDPRVPIMCWSCATYTIQSIERILSDKPLFPLCRDLDCLDSLTRPAAAHWT
VASVYVQGHFCKILFASLVNDSHEELPCILDIMFHLVLGLVLAFFALQCODFSGIS
LGTGDLHIHLVMTAAHITIIILTSCTEENQDQENPCEESAVIALKTKLHYTGSA
LKEIPSGHMLWRSVRGIMPFLKCSALFHYLVNVPDPDIQVPGTSHPELUCSYLSL
PNNLICLPQENSEIIMNSLIESWRNSEVKRYLEGERDARIYPRESNKLNLHPEDYSSL
INQASNSCPKSGGDKSRAPILCLVCGSLLCSQSYCCQTELEGEDEVGACATATYSCGS
LWGLIFLRVRECQVILFAGTKGCFYSPYLDYGETDQGLRGRNPLHLUCKERPKFIQK
LMHQSVTYEIGHAQEANQTLVGDIMQHL"
BASE COUNT 1618 a 1015 c 1193 g 1442 t
ORIGIN
Query Match 21.9%; Score 1378.6; DB 9; Length 5268;
Best Local Similarity 56.4%; Pred. No. 0;
Matches 2945; Conservative 0; Mismatches 2169; Indels 105; Gaps 16;
QY 780 TGGGATCAGCAAGTTGATTTTATATCTCTTCTTGCATCATTTTGCACAAATTTGGTGCA 839
Db 82 TGGCTGCAAGCACTGACCTCCTCAGAGAAGTGTACAGCAATTTAGCCCACTATGTACCC 141
QY 840 GAAATTTACTTTGCTGAAATGGACCCAGACTTGGAAAGCAGGAGGAGAAAGTGTACAAATG 899
Db 142 AAATCTACTCCAGGGGTCCCAACCCITTTCCACAGAAAGAGACATGCTGGCACAGCAT 201
QY 900 TCAATATTACTCTCCTCAGGAATGTTACTTATTTGGAGAAGATGCCAGATATTTGCTTAGAG 959
Db 202 GTTTTGTGGGAACCAATGTAATGTTACCTTTGTGTGAAGATCCTGCTGCAATTTGGATTTCCA 261
QY 960 AATTTGAAGCAC---AGTGGAGCATTTTCAGCTTTGCGGAGGGTTTTCAAAAGTGGAGAG 1016
Db 262 AAATTCAGCAAGCAAAACAAACCTTCTCATCTTTGTGTCGTGTTTAAAGATAGGAGAG 321
QY 1017 ACAACCTATTCTTCAGGGATTGTGCAATTCATCAACATGTGTACTCTGTATGACTGCTC 1076
Db 322 CTRACATATCTTCAGAGACTGTGCAGTTGATCCAACTTGTGTTTGTGATGAGAGTGC 381
QY 1077 TTCAGGACAGTGTTCATAAAATCATCGTTTACAGATGCATCTTCTACTGGAGAGGG 1136
Db 382 TTTTGGGAGTATTACAGAGATCATCGATAGATGACAACATCAGGAGGTGGAGGT 441
QY 1137 TTCTGTGACTGTGGAGACACAGAGCATGGAAACTGGCCCTTTTGTGTAATCATGAA 1196
Db 442 TTCTGTGACTGTGTGACTTGAAGCCCTGGAAGAGGGTCTCTTACTGTCAAAAGATGAA 501
QY 1197 CCTGGAAGACAGTACT--ATAAAGAGAAATTCACGCTGTCCTCTTGAATCAAGAGGTA 1253
Db 502 CTTAACACCTCTGAAATTTGAGGAAGAGAGGATCCTCTTGTTCATTTATCAGAAGATGG 561
QY 1254 ATTGTCCAAGCCAGGAAAAATATTCTTCAGTGATAAAATATGTCGTAGAAATGACTATA 1313
Db 562 ATAGCAAGAACTTATAACATTTTTCCTATTACGTTTCGTATGTCAGTAGAATATTAACC 621
QY 1314 TGGGAAGAGGAAAAAGAACTGCCCTCCCTGAACTCCAGATGAAGGAGAGAAAAATGAAAGATAC 1373
Db 622 TGGAAAAAAGAAAGTGAATTCGCCAGCAGATTTAGAGATGGTAGAGAGAGTACACACTAC 681
QY 1374 TATTGTCTCTTTTCAATGATGACACCAATTCATATGACCAGCTCATATACAGCCTACAA 1433
Db 682 TATTGTGCTGTTTAAATGATGAGGTTTACACCTTATGAACCAAGTATTATTACTCTTCAG 741
QY 1434 AGAGCTCTTCACTGTGAGCTCGCAGCGCCCGCCAGTTGTCATACACTGCCATTCGACAAAGAG 1493
Db 742 AAAGCTCTTAACTGTACAAAAAAGAGCTATTGGTTTTCACATACAGTAGATGTCAGAT 801
QY 1494 GGTCGTCGGGTGTTAAAGCGGGAGCTTATGCTGCTTCCGAGGAAGCAAGAAAGATATA 1553
Db 802 GGGCGTAGGCTGTGTCGATATGAGATTTTTCAGTATTGTCGAGCAACAAATCAGTAAT 861
QY 1554 AAGAGTCATTCAGAAAAATGCTCTCAACATCCACTTCATGTAGAGAAGTATTACATCAGAG 1613
Db 862 GTGAGAAATACCAGTAGACAGAC---AAAGCCACTCAAAGTTCAAAGTTTATGATTCGTC 918
```

QY 1614 ATTATGGCTCATCAGAAATTTGCTTGGCTTCTGTTTCCCTGGATGAACAAAATTTATGAGC 1673  
DB 919 ATTGTCGCACATCAGAAATTTGGTTTGAACCTTTTGGCTGGGAGTATTTATGGA 978  
QY 1674 TATTCAAGTGAATTTAGGAGATCTTTTGGCAAGCATGCCCTAGAGAAACCTGACTCG 1733  
DB 979 TATTTCAGATGCCCTTCGCCGATTTTATGTCAAAGTTGGTTTACAAAGGAGGCGGAGATG 1038  
QY 1734 GAGAAATCCCTGCTCATCAGAGGTTTAAATGCTTTGGGATGCAAGCTTTATAGAGTGCC 1793  
DB 1039 GAAACACTCTTCTCTAGTGGACAGACTGATGCTTATGATTTCCAAATTTATGGAAGGTC 1098  
QY 1794 CGTAAGATCCCTTCAATGATCTTCAGCAGTTTATTTATGGAGATGGAATACAAAAA 1853  
DB 1099 AGGAGTATATCATCAGTTGTTTCTCAGCAGCTGCTTATGGATTTGAAATACAGAA 1158  
QY 1854 CTCCTTCTATGGAATTTTGAAGTATTTAAACAACTGCAGAAAGATATATCAGTAT 1913  
DB 1159 CTATTTGCTGTCGATTTGCAAAAAATTTACCAGCAGTTGCAGAGATTTTATGGAGAT 1218  
QY 1914 GATCATCAGAGAGTATCTCTAATCTGACCTTTTCAGTTTCAGATGTTTACTGTTCTACT 1973  
DB 1219 GATCAGAGGAGCAGTGTGCTGACTGCTCTATCTGTCTCAGTTCTTCCAGCTTCTTCC 1278  
QY 1974 CTGGCTCGACATCTTATTGAGAGCAGAAATGTTATCTGTCATTTACTGAACTCTGCTA 2033  
DB 1279 CTGGCTCGAATGCTCATCAGAGAAACTTGTATGAGCAFTATCATTAAGACTTTTATG 1338  
QY 2034 GAAGTTTTT-----ACCTGAGTACTTGGACAGGAACAATAAATTCACCTTCAGGGTTAT 2087  
DB 1339 GATCATTTGAGACATCGAGATGCCAGGCGAGATTTTCAGTTTGAACGATACACTGCTTA 1398  
QY 2088 AGCCAGGACAAATTTGGAAGAGTATATGCAATATATGTGACCTTAAAGTATATCCTGATC 2147  
DB 1399 CAAGCCTTCAAAATTTAGGAGAGTACAGAGCCTTATTTTAGATCTCAAGTATGTTAAT 1458  
QY 2148 AGCAAAACCCACAATATGGACAGAAAGATTAAAGATGCACTTCCCTTGAAGTTTTCGATC 2207  
DB 1459 AGCAAAACCAATGAAATGGTTCAGATGAGCTGAGGAGAAAGTTCCCTAGAGGGTTTATG 1518  
QY 2208 TTTTGTGAAGATCTTACTGTATGACGAGGAAATGGAAGAAATCCGAAGACAGGTTGGGCA 2267  
DB 1519 TTTTGTGAATTTCTAAATGATGATGCGGAAATGATCCAAATTCACGTCAGATGAGCA 1578  
QY 2268 CACATTTGAAGTGGATCTCTGATTTGGAGGCTGCCATTTGCTATACAGATGCAATTTGAAGA 2327  
DB 1579 CATATTTGAATTTGAACACAGAGTGGAGGAGCCCTTCACACTACAAATGAATTAACAT 1638  
QY 2328 ATTTTACTCATGTTTCCAGAGTGTGCTGTTGATGAAGAACTCTTACTTGTGCTTAT 2387  
DB 1639 GTCATTTTCAATGATGACAGGAGTGTGCTGCTTCAGATGAAAAAGTGTAAATCGAAGCTT 1698  
QY 2388 AAGCAATGTCAAAAGCTGTATCAGTGCAGTACCAGTTTCATATCTAGTAGCAAGACA 2447  
DB 1699 AAGAAATGTCGCTGTACTGATGCGATGTCATCGGTGTTATCTGATGGTGAACAGCCA 1758  
QY 2448 GT----AGTACAAATCGTGTGGACATAGTTTGGAAACAAAGTCCCTACAGAGTATCTGAG 2504  
DB 1759 ATCACACTAAGCAATTTGGACATTCAGTGAACACTATCAGATCTGTTGTTCCCAAGAA 1818  
QY 2505 CTTGTAAGCATPATCTGCCACTCTCTAGGACCCTTGGCTGCTTCATGTACGTTTAAAGC 2564  
DB 1819 AAAGTTAGCATTCACCTCCCAAGTTCTCGCTTACTTGCAGGTTTACATGTTATTAAGC 1878  
QY 2565 AGGCTGGTGTCTTCAAGACTCCATGAATTTGCTGCTTTTGGAGGACTTTCAAGTAGAG 2624  
DB 1879 AAAAGTGAAGTGGCATATAAATTTCCAGAGCTCTACCTCTAAGTGAACCTTAGGCCACCC 1938  
QY 2625 GTACTAGTGAATATCTTTACGTTGCTGCTGTTGTTGGTCCAGGTTGTTGCTGAGATG 2684  
DB 1939 ATGTTGATGAACACCCCTCTAGATGCTGTTCTGTTGCTCCCAAGTACATGCGCGAATG 1998  
QY 2685 TGGCGAAGAAATGGACTGCTCTTATTAGCCAGGTTTATTACCAAGATGTTAAAGTGC 2744

DB 1999 TGGAGAAAGAAATGGGTTCTCTCTAGTAACACAGATTTATTACTACCAATATGTGAATGC 2058  
QY 2745 AGAGAAGAAATGATATGATAAAGATATCATCTGCTTCAGATTTGGTGCATCTTTTAATGGAT 2804  
DB 2059 AGACGTGAGATGTTTGCACAAAGGATGTAGTAATGCTTCAGACAGGTGTCTCCATGATGGAT 2118  
QY 2805 CCCAATAAGTCTTCTGTTACTTGTACTTCAGAGGATGAACATTTGCCAGGCTTTTAACAAG 2864  
DB 2119 CCAATATCATTTCTCTGATGATCATGCTCAGCCGCTTTGAACCTTTATCAGATTTTCACTACT 2178  
QY 2865 ACCATATCTACAAAA-----GACCAGSATTTGATTTAAACAATAT 2903  
DB 2179 CGAGACTATGAAAAAGATTTTAGTTCTGAGATTTACCATAAGGATGTTGTTCCAGCAGA 2238  
QY 2904 AATACACTAATAAGAAATGCTTCAGGTCTCTCATCTATATTGTTGGGTGAGGCTTTATGTA 2963  
DB 2239 AATACTCTAATAGAAAGATGCTATACCTCATTAATAATGCTTGTGTCGAGAGATTTAGT 2298  
QY 2964 CTGGAGTGGGAAATGTGACCAAGAGGTCACAATCAGAGAAATCATTCACCTGCTT 3023  
DB 2299 CTGGAGTGGACAGGTAATGCTACAGATGAATCAAGCAGAGATTTATCCATCAGTTG 2358  
QY 3024 TGCATTGAACCCATGCCACACAGTCCATTGTCACAAAATTTACCTCAGAAATGAAAAAT 3083  
DB 2359 AGTATCAAGCCTATGGCTCATAGTAATTTGGTAAAGTCTTTACCTGAAGATGAGAAACA 2418  
QY 3084 GAACTGCTTAGAGATGCTATAAACAAGTGGCCACATTTTAAAGAAACCAAGCTGTATCA 3143  
DB 2419 GAGACTGCGATGGAGAGTATTCGAGCAGTTGCCCATTTCAAGAAACCTGGATTAACA 2478  
QY 3144 GGCCATGAGTGTATGAACTAAAGATGAATCACTGAAAGACTTCAATATGACTTTTAT 3203  
DB 2479 GGACGAGGATGATGAACTGAAACAGAAATGTGCCAAAGAGTTCAACTGTTATTTCTAT 2538  
QY 3204 CATTTACCAACCCAGCATAGCAGGCTGAACATATCAGAGAAAGAGGAGAAACAA 3263  
DB 2539 CACTTTTCAAGGCGAAGACAGTCCAAAGGCGAAGAACGCGAAATTTGAAAGACAA 2598  
QY 3264 GAAAAACAAGATGAAGCATTTGCCGCCACCCACCTCTCGAATTTCTGCCCTGCTTCAGC 3323  
DB 2599 AATAGAGAGATACAGACTCCACCTCCGCTGTTGCCCTCCATTTGCCCTCTGTTGCA 2658  
QY 3324 AAAGTGATTAACCTTTCTCAACTGTGATATCATGATGTACATTTCTCAGGACCGTATTGAG 3383  
DB 2659 AGCCTGGTTAACAATTTTGCAGTCAGATGTCATGTTGTGTCATCATGGAAACAATTTCTG 2718  
QY 3384 CGGCAANTAGACACAGATCTTAACCTTGTGGACCGAAGGAGTGTCCAATGGCTTTTCAT 3443  
DB 2719 TGGGCTGTGGAACATAATTTGGATATGCTGCTGTCAGAGTCCATGTCGAAAGGTTTACAT 2778  
QY 3444 ATTTCTGGCATTTGGGTTTACTAGAGAGAAAGCAACAGCTTCAAAAAGCTCCTGAAG---AA 3500  
DB 2779 TTAATTTGATGCGCACTACAGAGAAAGAAACAACATTTAGAGAAATGTCAGGAGAGCAT 2838  
QY 3501 GAAGTAAACATTTGACTTTTATCATAGGCTTCAAGATTTGGAAGTTTCAGCCATGAAT--- 3557  
DB 2839 GTAGTAACATTTTACCTTCACTCAGAAGATATCAAAACCTTGTGAAGCGCAAAAATTTCT 2898  
QY 3558 -----ATACAATGCTTTTGGAAAACATCAAGGAATTTCCCAAGTTAGAGGCCAGAG 3611  
DB 2899 CCTAGCATACTAGCTATGCTGGAAACACTACAAAAATGCTCCCTACCTAGAAAGTCCACAA 2958  
QY 3612 GACATGATAACCTGGATCTTACAGATGTTTGACACAGTGAAGCGATTAGAGAAAATCT 3671  
DB 2959 GACATGATTCGGTGGATATTTGAAGACTTTTAAATGCTTTTAAAGATAGGGAGAGTTCA 3018  
QY 3672 TGTTTAAATTTAGCAACACATCAGATCGGAATCTTATTAAAGATGATCAGATTTACTCAT 3731  
DB 3019 CCTACCAGTC-----CGTGCAGAGACAGAGGAAACCATTAATGGAAGAGTTCAAGG 3072  
QY 3732 GATAAAGAAAAAGCAGAAACGAAAAAGCTGAAGCTGTAGGCTACATCGCCCAAG 3791  
DB 3732 GATAAAGAAAAAGCAGAAACGAAAAAGCTGAAGCTGTAGGCTACATCGCCCAAG 3791



Db 3073 GACAAAGACAAAGCTGAGAGGAGAGAAAGACAGAGATTCCAGACTGCCACAGAAAAG 3132  
QY 3792 ATCATGGCTCAGATGCTCGCTTACAGAAAAAATCTCATGAAACTCATAAACTCATGAT 3851  
Db 3133 ATCATGGCTCAGATGCTGAAATGCAAGCGCATTTTATTGATGAAACAAAGAACTCTTT 3192  
QY 3852 GACATACATCAGAAATGCTGGGAAGAAGATTCATATGAGGAAGAGACACCCCA 3911  
Db 3193 CAGCAGACATTTAGAACTGGATGCTCAACTCTGCTGTTCTTGATCA-----TAGCCCT 3246  
QY 3912 GCAGTCAGTGACTACTTCTAGAAATGCTTTGGGTCTTAAACGGGGTCCATCTGTTACTGAA 3971  
Db 3247 GTGGCTTCAGATATGACACTTACAGCACTGGGCCCCGCACAACTCAGGTCTCTGACAA 3306  
QY 3972 AAGGAGTCTGACGTGATCCTTTGCCAAGAAACAGAGGTGAAATAGAAAATAAT 4031  
Db 3307 AGACAATTCGTTACATGATATTTGTCTAAGAGGAGCAAGTCTTAAAGTGGAAAGCAGG 3366  
QY 4032 GCCATGGTATTATCGGCTGTGCCAGAAATCTACTGCCTTAACCCAGCACAGGGGAAA 4091  
Db 3367 GCAATGGTCTTGGCAGCATTTGTTCAGAGATCAACTGTATTATCAAAAACAGAAATAA 3426  
QY 4092 CCCATAGAACTCTCAGGAGAAGCCCTAGACCCCACTTTTTCATGATCCAGACTTGGCATAT 4151  
Db 3427 TTATTCAAGATCCAGAAAA---TATGATCCATTAATTCATGCACCCCTGATCTGTGT 3483  
QY 4152 GGAATTTATACAGGAAGCTGTGTCATGTAATGCACGAGTGTGTCGGCAAGATATTTT 4211  
Db 3484 GGAACACACACTAGTAGCTGTGGGCACATTAATGATGCCCATTTGTTGGCAAGGATATTT 3543  
QY 4212 GAAGCTGTACAGCTCTCAGACGCCATTCATGTTGACCTTT-----T 4259  
Db 3544 GATTCGGTTCAAGCTTAAAGAACAGCGAAGGCAACAGAGATTACGCTTACATAGAGCTAT 3603  
QY 4260 GACTTGGAAAGTGGAGAATATCTTGCCCTCTTGGCAATCTCTGCAATCTGTCATCTGATC 4319  
Db 3604 GATGAGAAACGGAGAATCTCTTGCCCCCTTTGGAATGCTGTGATGATATCTGTATT 3663  
QY 4320 CCCATATTCTCTTGCACCTCAAAAGATAAACAGTGGAGAATGCAGATGCTTGTCTCAA 4379  
Db 3664 CC-----TCTGCTCTCTCCAGAAATATTTTAAACACAGGTTAAATTTTTCAGAC 3717  
QY 4380 CTTTGTACCTCGGACGGTGGATACAGACTGTTCTGCCAGAAATATCAGGTTATATATA 4439  
Db 3718 CAACCAAAATCTGACTCAGTGGATTTAGAACAAATATCTCAGCAATAAAGACATTTACAGTT 3777  
QY 4440 AGACATGCTTAAAGGAGAAACCAATTCCTATTCTTTTAAATCAAGGAATGGAGATTCT 4499  
Db 3778 CTTAGGAAGAAGAAGTACTCTTAATAATGCCCTACAAAGAAATTCAGAAATGTGGAT 3837  
QY 4500 ACTTTGGAGTTCCATCTCCTGAGTTTGGCGTTGAGTCTTCGATTAAATATTCAAAT 4559  
Db 3838 GAATTACAGCTCCCTGAAGGTTACGGCTGATTTTCGTCCTTAAGATCCCTTATTCTGAG 3897  
QY 4560 AGCATCAAGAAATGTTATTCCTTTGGCAACAAATTTATAGAAATGTGTAAGT 4619  
Db 3898 AGCATAAAAGAAATGCTAACAGACATTTGGAACGTCTACCTACAAAGTGGGACTAAAAGTT 3957  
QY 4620 CCACCTGTGAAGGATCCTCGAGTCCCATCTGACCTGGAGCACTGCGCTTTCACT 4679  
Db 3958 CATCCCAATTAAGAGGATCCTCGTGTCCCTAATATGTTGGGTAGCTGCGCGTACAC 4017  
QY 4680 ATCCAGGCAATTTGAAATCTATTGGGAGATGAAGGAAACCTCTGTTTGGAGCACTTCAA 4739  
Db 4018 ATCCAAAGCATAGAAGAAATTTTGATGATGAAGATAAACCATTGTTTGGTCTTTACCT 4077  
QY 4740 AATAGGCATCAATAGGTTGTAAGCATTAATGTCAGTTTGCAGTTGTCACAGAGGATTAAC 4799  
Db 4078 TCCAGACTGGATGACTGCTTAGTTCATGACGAGATTTGCCGCACACACTGGACAGTG 4137  
QY 4800 TGTCTCAGGCTCTGATACAGAAACATCTGGTTCGCTTCTATCATGTTGTTCTTCTCAAC 4859  
Db 4138 GCATCAGTTTTCAGTGGTGAAGGACATTTTGTAAACTTTTGTCACTCACTGGTGGCTAAT 4197

QY 4860 ATAAATCAGAAGATACACATGCTTCTGTCTATATAGATCTGTTTTCATGTTTGTGGGT 4919  
Db 4198 GACAGCCATGAGGAACCTTCCATGATATATAGATATTCATGTTTTCATTTATTGTGGGC 4257  
QY 4920 GCTGTGTTAGCATTCCTTCTGTTATTTGGGATGACCCCTGTTGATCTCGACCTTCTTCA 4979  
Db 4258 TTGGTGTTCATTTCTTCCGCTTCAGTG-----TCAGATTTTTCAGGG 4302  
QY 4980 GTTAGTTCTTCTATAACCACTTTATCTTCTTCCATTTTGATCACATGGCACATGCTT 5039  
Db 4303 ATCAGCCCTGGCACTGGAGACTTTCACATTTTCCATCTGGTTACTATGSCACACATATA 4362  
QY 5040 CAGATACTACTTACAGTAGACACAGGCC---TACCCCTTGTCTCAGGTTCAAGAACACAGT 5096  
Db 4363 CAGATCTTACTTCTCATGTACAGAGAGAAATGGCATGGATCAAGAAATCCCTTGT 4422  
QY 5097 GAAGAGCTCATTCCTCGCATCTTCTTCTTCAGAAAAATTTCTCAATATACAGTGGCTCC 5156  
Db 4423 GAAGAAGAACAGCAGTTCTTGTGTTATAAACACTTCACCACTATACGGGAAGTGCC 4482  
QY 5157 ATTGGGTGTGATATTCCTGCTGCTTATTTGGGTCTTCACTGAAGAAATGGCATCAACCT 5216  
Db 4483 TTGAAGAAATACCATCGGCTGGCATCTGTGGAGGAGTGTACAGAGCTGGAATCATGCT 4542  
QY 5217 TATCTTCGCTGTGCTGATTTTTCACATTAATTTACTTGGGGTAACCTCCGCTCAGGAA 5276  
Db 4543 TTCTGAAGTGTCTGCTTTATTTTTCATTAATAATGAGTTCCTTCCCCACCCGAC 4602  
QY 5277 CTGATACCAATTTCTGCAGAGAGAGTACAGTGCATCTGTAGTATCTATCTTACT 5336  
Db 4603 ATTCA---AGTTCTCGAACAAGCCATTTTGAACATTTATAGTATCTTTCCCTACCA 4659  
QY 5337 ACAAAATTTGTTCTGCTCTTCCAGGAATATTTGGGATCTGTAAAGCCCTTGTCCAGAGG 5396  
Db 4660 AACACCTCATTTGCCCTTTTCAAGAAATAGTGAATGAATTAATCTGATGAAAGT 4719  
QY 5397 TGSTGTGCAGATCTGCTTACTAACTGTTTGAAGCAAAAAACACCGTGGTCAAGTAC 5456  
Db 4720 TGSTGCGGTAAACAGTGAATTTAAAGATATCTAGAAGGTGAAGAGAGTCTATAAGAT 4779  
QY 5457 CTTAGAAAAAGAAATAGTTGATAGAGTTCCTGATGACTATAGTGGCTCTCTGAATCAA 5516  
Db 4780 CCAAGAGAATCTAACAAATTAATAACCTTCCAGAGGATTTACAGCACTCATTAATAA 4839  
QY 5517 GCTTCTCATTTCAAGTCCCGCTGTGCATGATGACGAAAGCACTCTGCTCTGCTG 5576  
Db 4840 GCATCCAAATTTCTGCTCCCGAATCAGGTGGTGAAGAGCAGAGCCCCAACTCTGTGC 4899  
QY 5577 CTTTCTGTGGGGTATATCTTCTCAGAAACATTTTGTCTCCAGGAAATTTGTGAACGG 5636  
Db 4900 CTTGTGTGCGGATCTCTGCTGTCTCCAGAGTTACTGCTGCCAGACTGAATGGAAGG 4959  
QY 5637 GAAGAGTTGGAGTTGCATTTTTCAGCACTTCTCCTGTGGAGCGGAGTCTGCAATTTTC 5696  
Db 4960 GAGGATAGGAGCTGCACAGCTCACACCTACTCTCTGCTCTGGAGTGGGCACTCTTC 5019  
QY 5697 CTAAAAATCAGAAATGCCAGTGTCTTGGTTGAAGTAAAGCCAGAGGCTGTGCCCTAT 5756  
Db 5020 CTGAGAGTACGGGAATGTCAAGTGTCTATTTTGTGCAAAACCAAGGCTGTTTTAT 5079  
QY 5757 CCAGCTCTTACTTGGATGAATATGGAAACAGACCTTGGCTTGAAGAGGCAACCC 5816  
Db 5080 TCTCCCTCTTACCTTGTGATGATGGGAGACCCACAGGACTCAGACGGGGAATCT 5139  
QY 5817 CTTTCAATTTCTCGTAGCGGTATCGGAAGCTTCCATTTGGTCTGCCAACAACTGCAAT 5876  
Db 5140 TTACATTTATGCAAGAGCGATTCAGAAGATTTCAGAAGCTCTGGCAACCAACAGTGTCT 5199  
QY 5877 ATGAGAGAAATTCGTAGAGCCCAAGACACTAATCAGATGTTATTTGGATTTCACTGGCA 5935  
Db 5200 ACAGAGGAAATTTGGACATGCACAGAAAGCCAAATCAGACACTGGTTGGATTTGACTGGCA 5258







Db 3356 TGGGAAGCAGGCGATGGTCTTGGCAGCGTTTGTTCAGAGTCAACGGTTCTGTCAAAG 3415  
Qy 4081 ACAGGGGAAACCCATAGAACTCTCAGGAGAACCCCTAGACCCACTTTTTCATGATGATCAG 4140  
Db 3416 ACAGGACGAAACCATCGGACCCAGAGAAAA---TATGATCCATTAATCATGACACCCG 3472  
Qy 4141 ACTTGGCATATGGAACCTATACAGGAAGCTGTGCTGATGTAATGCACGACGCTGCTGGC 4200  
Db 3473 ATCTGTCTTGGGACACACTGGCAGCTGTGGCAGGTTATGCATCCCATGTTGTGGC 3532  
Qy 4201 AGAAGTATTTGAAGCTGTAC-----AGCTGAGCTCTCAGCAGCGCATTCATGTTG 4251  
Db 3533 AAAGGTATTTGATTCCTTCAAGCCAGGAGCAGCAGGAGCGGCTGCGCTTGC 3592  
Qy 4252 ACCTT---TTTGACTTGGAAAGTGGAGATATCTTTGCCCTCTTTGCAAACTCTCTGTGCA 4308  
Db 3593 ACATGATACGATGTAGAGATGGGAGTTCCCTGCTGCCGCTCTGTGAGTGCCTGAGCA 3652  
Qy 4309 ATACTGTGATCCCATTAATCTTCTTGGAACTCAAAAGATAAAGAGTGAAGATGAGATG 4368  
Db 3653 ACACGGTGATCCCC-----CTGCTGCTTCTCCAGGAGCATCTCAGCAGGAGGTTAA 3706  
Qy 4369 CTCTTGCTCAACTTTTGACCTTGCACGGTGGATACAGACTGTTCTGGCCAGATATCAG 4428  
Db 3707 ATTTTTCAGACCAACCATGCTGGCAGAGTGGACAGTGGAGAGCAGTAACACAGCAGATAAGG 3766  
Qy 4429 GTTATAATATAGCATGCTAAAGGAGAAACCCCAATTCCTATTTTCTTAATCAAGAA 4488  
Db 3767 TGGTCCAGATCTGAGGAGAAAGACATGCTGTGACACGCTCTTTCAGAGGACACAG 3826  
Qy 4489 TGGGAGATCTTACTTTGGAGTTCATATCCATCTGAGATTTTGGGCTTGAGTCTTCGATTA 4548  
Db 3827 AAGCATATATATACCGATCCCGAAGGCTTCAGGCTGATTTTATCTAGTAACC 3886  
Qy 4549 AATATCAATAGCATCAAGAAATGGTTATCTCTTTGCCACACAAATTTATAGAAATG 4608  
Db 3887 CATATCTGATAGCATAAAGAAATGTTAAGACATTTGGAACGGCTCTTACAAGGTGG 3946  
Qy 4609 GATTGAAGTGCACATGATGAAGGGATCTCGAGTCCCATGCTGACCTGCAGCACCT 4668  
Db 3947 GACTGAAGTTCATCTTAATGAGGTGACCCCGTGTGCCATCTCTGTGCTGGGGACCT 4006  
Qy 4669 GCGTTTTCATATCCAGCAATGAATAATCTATTGGGAGATGAAGAAACCTCTGTTG 4728  
Db 4007 GTGCATACACCATCCAGAGCATAGAAAGATTTTGAGTGATCAGGAGAGCCTGTTTGTG 4066  
Qy 4729 GAGCACTTCAAAATAGGACGATATGCTGCTGAAGCATTAATGCACTTTGCACTTGCAC 4788  
Db 4067 GACCTCTGCTGTAGACTGGACGACTGCTCAGGCTGTTAACACGGTTTGCAGCAGCAC 4126  
Qy 4789 AGAGGATTACCTGTCTCAGTCTGATACAGAAACATCTGGTTCCTCTCTATCAGTTG 4848  
Db 4127 ATTGACAGTGGCGTACTCTGTGGTACAGGACACTCTGTAACACTCTTGCATCT 4186  
Qy 4849 TTCTTCTTACATAAAATCAGAAGATACACCATGCTCTCTGCTATAGATCTGTTTCATG 4908  
Db 4187 TGGTCCCTAGTGACAGCTATGAAGACCTCCCGTGCACTACTAGACATCGACATGTTTCACT 4246  
Qy 4909 TTTTGGTGGTGTGTGTAGCATTTCCCATCTCTGATTTGGATGACCTGTTGATCTGCG 4968  
Db 4247 TGCTGTGGGCTGTGTGCTGCTTTCCAGGCTCTGCAGTG-----TCAGG 4291  
Qy 4969 AGCCTTCTTCACTGATGTTCTTCTTATACACCTTTATCTCTTCCATTTGATCACCATGG 5028  
Db 4292 ATTTTTCAGGAGACGCTGGCCACTGGGACCTGCACATCTTCCACTTGGTTACATGG 4351  
Qy 5029 CACACATCTCTCAGATACTACTTACAGTAGACACAGCGCTACG---CCTTGTCTCAGGTTTC 5085  
Db 4352 CACACATCTCAGATCTTACTTACTCTATCTACAGAGAGATGGCATGGATCAAGACA 4411  
Qy 5086 AAGAAGACAGTGAAGAGCTCATTCGCCGATCTCTTCTTCTTGTGAGAAATTTCTCAATATA 5145

Db 4412 ATCCCACTGGGAAGAAGAACTGGCCATTTCTCTCTTTGCACAAAACACTTTCACAGTATA 4471  
Qy 5146 CAAAGTGGCTCCATTTGGGTGTGATATTCCTGGCTGGTATTTTGTGGTCTCACTGAAGAAATG 5205  
Db 4472 CTGGAAGTCCCTTGAAGAAGACCCCTCGGCTGGCACCTGTGGAGAGCGCTCGGCGCG 4531  
Qy 5206 GCATCACCCCTTATCTTCCGCTGTGCTGATGTTTTCCTACTATTTTACCTTGGGTACTC 5265  
Db 4532 CCATCATGCTTTCCTCAAGTGTCTGCTTTGTTTTCCTACTATTTAAATGGAGTCCCGG 4591  
Qy 5266 CCCTGAGGAACCTGCATACCAATTCCTGCAAGAGAGATACAGTGCACCTCTAGCTATC 5325  
Db 4592 CCCTCCAGACCTTCA---AGTTCTGGACAGACCATTTTGAACACTTATGTAACTACC 4648  
Qy 5326 TATCTTTACTACAAATTTTCTCTCTCTTCCAGGAATATTTGGGATCTGTAAAGCCCT 5385  
Db 4649 TTTCCCTACCAACCAACCTCATTCACCTTTTTCAGAAAACAGTGCATCATGAACCTCC 4708  
Qy 5386 TGCTCCAGGTGGTGTGCAGATCCTTCTTACTAACTGTTTGAAGCAAAAACACCG 5445  
Db 4709 TGATTGAAAGTTGGTGCCAGAACAGTGAAGTTAAACGGTATCTAAATGGCAGAGAGAG 4768  
Qy 5446 TGGTCAAGTACCTTACAAAAAGAAATAGTTTGTATAGAGCTTCTCTGATGACTATAGCTGCC 5505  
Db 4769 CATAAGCTACCCAGAGAGGCTAACAACTGATAGACCTTCCAGAGGATTTACAGAGCC 4828  
Qy 5506 TCCTGAATCAAGCTTCTCATTTTCAAGTGCACGGTCTGCAGATGATGAGCGAAAGCATC 5565  
Db 4829 TCATTAACCAAGCATCCAATTTCTCTGCTGCCCAAAATCAGGTGGCGCACAGCAGAGCTC 4888  
Qy 5566 CTGCTCTCTGCTTCTTCTGCTGGGCTTACTATGTTCTCAGAACATTTTGTGCGAGGAAA 5625  
Db 4889 CTACTCTGTGCTGCTGCTGCTGGGAGTCTCTCTGCTCTCAGAGTCTTCTGCTGCAAGCTG 4948  
Qy 5626 TTGTGAACGGGAAGAGTTGGAGCTTGGATTTTTCAGGCACCTTCTACTGCTGGAGCCGAG 5685  
Db 4949 AGCTGAGGCTGAGGAGCTGCGAGCTGCACAGCACACACTTACTCTCTCGGCTCCGGGG 5008  
Qy 5686 TCTGCTATTTTCTTAAATAACAGAGATGCGAGTGGTCTGCTTGAAGGTTAAAGCCAGAG 5745  
Db 5009 CCGGCATCTCTGAGAGTGGGGAATGTAGTGTCTATTTTGTGCGCAAAACCAAG 5068  
Qy 5746 GCTGTGCTTATCCAGCTCTCTTACTTGGATGAATATGAGAAACAGACCTTGGCTTGAAGA 5805  
Db 5069 GATGTTTTTATCTCTCTTCTTCTGAGTGTGAGAGAGCCGACAGGACTCTAGAC 5128  
Qy 5806 GGGCAACCCCTTCTATTTATCTGAGGGTATCGAGCTCCCATTTTGTCTGCGCAAC 5865  
Db 5129 GAGGAATCTTTACATTTTATGCAAGAGCGGTTTGAAGATCCAGAGCTCTGCGAGC 5188  
Qy 5866 AACACTGCTATTAAGAGATTTGCTAGGAGCCAGAGACTAATCAGATGTTATTTGGAT 5925  
Db 5189 AGCATAGTATCAGAGAGATCGGACACGCGCAGGAGGCTAACAGACCTTGTCTCGAA 5248  
Qy 5926 TCAACTGGCAG 5936  
Db 5249 TTGACTGGCAG 5259

## RESULT 12

AR030785  
LOCUS 1001 bp DNA linear PAT 29-SEP-1999  
DEFINITION Sequence 2 from patent US 5861312.  
ACCESSION AR030785  
VERSION AR030785.1 GI:5943999  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 1001)  
AUTHORS Varsnavsky,A. and Kwon,Y.tae.  
TITLE Nucleic acid encoding mammalian UBRI  
JOURNAL Patent: US 5861312-A 2 19-JAN-1999;

FEATURES  
source Location/Qualifiers  
1. .1001  
/organism="unknown"  
BASE COUNT 363 a 186 c 205 g 247 t  
ORIGIN

Query Match 15.8%; Score 996.2; DB 6; Length 1001;  
Best Local Similarity 99.7%; Pred. No. 9.4e-241;  
Matches 998; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 2799 ATGGATCCCAATAAGTCTTGTACTGGTACTTTCAGAGGTATGAACCTTGCAGGCGCTTTT 2858  
Db 1 ATGGATCCCAACAAGTCTTGTACTGGTACTTTCAGAGGTATGAACCTTGCAGGCGCTTTT 60  
QY 2859 AACAAGACCATATCTCAAAAGACAGGATTTGATTAACAATAATATACACTAATAGAA 2918  
Db 61 AACAAGACCATATCTCAAAAGACAGGATTTGATTAACAATAATATACACTAATAGAA 120  
QY 2919 GAAATGCTTCAGTCTCATCTATATTGTGGGTGAGGTTATGTACCTGGAGTGGGAAT 2978  
Db 121 GAAATGCTTCAGTCTCATCTATATTGTGGGTGAGGTTATGTACCTGGAGTGGGAAT 180  
QY 2979 GTGACCAAAAGAGAGGTCAAAATGAGAGAAATCATTCACCTTGCTTTGCATTGAACCCATG 3038  
Db 181 GTGACCAAAAGAGGTCAAAATGAGAGAAATCATTCACCTTGCTTTGCATTGAACCCATG 240  
QY 3039 CCACACAGTGCCTATGCCAAAATTTACCTGAGAAATGAATAATGAACCTGGCTTAGAG 3098  
Db 241 CCACACAGTGCCTATGCCAAAATTTACCTGAGAAATGAATAATGAACCTGGCTTAGAG 300  
QY 3099 AATGTCATTAACAAGTGGCCACATTTAAGAAACCAGGTGTATCAGGCCATGGAGTTTAT 3158  
Db 301 AATGTCATTAACAAGTGGCCACATTTAAGAAACCAGGTGTATCAGGCCATGGAGTTTAT 360  
QY 3159 GAACTAAAAGATCAATCACTGAAAGACTTCAATATGTACTTTTATCATTTACTCCAAAACC 3218  
Db 361 GAACTAAAAGATCAATCACTGAAAGACTTCAATATGTACTTTTATCATTTACTCCAAAACC 420  
QY 3219 CAGCATAGCAAGGCTGAACATATGCAGAGAAAGGAGAGAAACAAAGAAACAAAGATGAA 3278  
Db 421 CAGCATAGCAAGGCTGAACATATGCAGAGAAAGGAGAGAAACAAAGAAACAAAGATGAA 480  
QY 3279 GCATTCGGCCACACACCTCTCTGAATTCCTGCTGCTTTTCAGCAAAAGTATTAACCTT 3338  
Db 481 GCATTCGGCCACACACCTCTCTGAATTCCTGCTGCTTTTCAGCAAAAGTATTAACCTT 540  
QY 3339 CTCACACTGTATATCATGTATCATCTCAGGACCGTATTTGAGCGGGCAATAGACACA 3398  
Db 541 CTCACACTGTATATCATGTATCATCTCAGGACCGTATTTGAGCGGGCAATAGACACA 600  
QY 3399 GATTCCTAACTTGTGGACCGAGGATGCTCCAAATGGCTTTTCATATTCCTGGCATTTGGT 3458  
Db 601 GATTCCTAACTTGTGGACCGAGGATGCTCCAAATGGCTTTTCATATTCCTGGCATTTGGT 660  
QY 3459 TTACTAGAAGAGAGAACACAGCTTCAAAAAGCTCCTGAAGAAGAAAGTAAACATTTGACTTT 3518  
Db 661 TTACTAGAAGAGAGAACACAGCTTCAAAAAGCTCCTGAAGAAGAAAGTAAACATTTGACTTT 720  
QY 3519 TATCATAAAGCTTCAAGATTTGGGAAGTTTCAGCCATGAATATACAAATGCTTTTGGAAAAA 3578  
Db 721 TATCATAAAGCTTCAAGATTTGGGAAGTTTCAGCCATGAATATACAAATGCTTTTGGAAAAA 780  
QY 3579 CTCAAAGGAATTCCTCCAGTTAGAGCCCAAGGACATGATAACGTGGATCTCAGATG 3638  
Db 781 CTCAAAGGAATTCCTCCAGTTAGAGCCCAAGGACATGATAACGTGGATCTCAGATG 840  
QY 3639 TTTGACACAGTGAAGCGATTGAAGAAAAATCTTGTGTTAATTTGTAGCAACCATCATCAGGA 3698  
Db 841 TTTGACACAGTGAAGCGATTGAAGAAAAATCTTGTGTTAATTTGTAGCAACCATCATCAGGA 900  
QY 3699 TCGGAATCTATTAAAGATGATGAGATTACTCATGATAAAGAAAAAGCAGAACGAAAAAGA 3758  
Db 901 TCGGAATCTATTAAAGATGATGAGATTACTCATGATAAAGAAAAAGCAGAACGAAAAAGA 960

QY 3759 AAAGCTGAAGCTGTAGCTACATCCCGAAGATCATGGC 3799  
Db 961 AAAGCTGAAGCTGTAGCTTTCATCCCGAAGATCATGGC 1001  
RESULT 13  
ARI21464  
LOCUS ARI21464 1001 bp DNA linear PAT 16-MAY-2001  
DEFINITION Sequence 2 from patent US 6159732.  
ACCESSION ARI21464  
VERSION ARI21464.1 GI:14105040  
KEYWORDS  
SOURCE unknown.  
ORGANISM unknown.  
REFERENCE 1 (bases 1 to 1001)  
AUTHORS Varshavsky, A. and Kwon, Y. Tae.  
TITLE Nucleic acid encoding mammalian Ubr1  
JOURNAL Patent: US 6159732-A 2 12-DEC-2000;  
FEATURES Location/Qualifiers  
1. .1001  
source /organism="unknown"  
BASE COUNT 363 a 186 c 205 g 247 t  
ORIGIN  
Query Match 15.8%; Score 996.2; DB 6; Length 1001;  
Best Local Similarity 99.7%; Pred. No. 9.4e-241;  
Matches 998; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 2799 ATGGATCCCAATAAGTCTTGTACTGGTACTTTCAGAGGTATGAACCTTGCAGGCGCTTTT 2858  
Db 1 ATGGATCCCAACAAGTCTTGTACTGGTACTTTCAGAGGTATGAACCTTGCAGGCGCTTTT 60  
QY 2859 AACAGAGCCATATCTACAAAAGACAGGATTTGATTAACAATAATATACACTAATAGAA 2918  
Db 61 AACAGAGCCATATCTACAAAAGACAGGATTTGATTAACAATAATATACACTAATAGAA 120  
QY 2919 GAAATGCTTCAGTCTCATCTATATTGTGGGTGAGGCTTATGTACCTGGAGTGGGAAT 2978  
Db 121 GAAATGCTTCAGTCTCATCTATATTGTGGGTGAGGCTTATGTACCTGGAGTGGGAAT 180  
QY 2979 GTGACCAAAAGAGGTCAAAATGAGAGAAATCATTCACCTTGCTTTGCATTGAACCCATG 3038  
Db 181 GTGACCAAAAGAGGTCAAAATGAGAGAAATCATTCACCTTGCTTTGCATTGAACCCATG 240  
QY 3039 CCACACAGTGCCTATGCCAAAATTTACCTGAGAAATGAATAATGAACCTGGCTTAGAG 3098  
Db 241 CCACACAGTGCCTATGCCAAAATTTACCTGAGAAATGAATAATGAACCTGGCTTAGAG 300  
QY 3099 AATGTCATTAACAAGTGGCCACATTTAAGAAACCAGGTGTATCAGGCCATGGAGTTTAT 3158  
Db 301 AATGTCATTAACAAGTGGCCACATTTAAGAAACCAGGTGTATCAGGCCATGGAGTTTAT 360  
QY 3159 GAACTAAAAGATGAATCACTGAAAGACTTCAATATGTACTTTTATCATTTACTCCAAAACC 3218  
Db 361 GAACTAAAAGATGAATCACTGAAAGACTTCAATATGTACTTTTATCATTTACTCCAAAACC 420  
QY 3219 CAGCATAGCAAGGCTGAACATATGCAGAGAAAGGAGAGAAACAAAGAAACAAAGATGAA 3278  
Db 421 CAGCATAGCAAGGCTGAACATATGCAGAGAAAGGAGAGAAACAAAGAAACAAAGATGAA 480  
QY 3279 GCATTCGGCCACACACCTCTCTGAATTCCTGCTGCTTTTCAGCAAAAGTATTAACCTT 3338  
Db 481 GCATTCGGCCACACACCTCTCTGAATTCCTGCTGCTTTTCAGCAAAAGTATTAACCTT 540  
QY 3339 CTCACACTGTATATCATGTATCATCTCAGGACCGTATTTGAGCGGGCAATAGACACA 3398  
Db 541 CTCACACTGTATATCATGTATCATCTCAGGACCGTATTTGAGCGGGCAATAGACACA 600  
QY 3399 GATTCCTAACTTGTGGACCGAGGATGCTCCAAATGGCTTTTCATATTCCTGGCATTTGGT 3458  
Db 601 GATTCCTAACTTGTGGACCGAGGATGCTCCAAATGGCTTTTCATATTCCTGGCATTTGGT 660  
QY 3459 TTACTAGAAGAGAGAACACAGCTTCAAAAAGCTCCTGAAGAAGAAAGTAAACATTTGACTTT 3518  
Db 661 TTACTAGAAGAGAGAACACAGCTTCAAAAAGCTCCTGAAGAAGAAAGTAAACATTTGACTTT 720  
QY 3519 TATCATAAAGCTTCAAGATTTGGGAAGTTTCAGCCATGAATATACAAATGCTTTTGGAAAAA 3578  
Db 721 TATCATAAAGCTTCAAGATTTGGGAAGTTTCAGCCATGAATATACAAATGCTTTTGGAAAAA 780  
QY 3579 CTCAAAGGAATTCCTCCAGTTAGAGCCCAAGGACATGATAACGTGGATCTCAGATG 3638  
Db 781 CTCAAAGGAATTCCTCCAGTTAGAGCCCAAGGACATGATAACGTGGATCTCAGATG 840  
QY 3639 TTTGACACAGTGAAGCGATTGAAGAAAAATCTTGTGTTAATTTGTAGCAACCATCATCAGGA 3698  
Db 841 TTTGACACAGTGAAGCGATTGAAGAAAAATCTTGTGTTAATTTGTAGCAACCATCATCAGGA 900  
QY 3699 TCGGAATCTATTAAAGATGATGAGATTACTCATGATAAAGAAAAAGCAGAACGAAAAAGA 3758  
Db 901 TCGGAATCTATTAAAGATGATGAGATTACTCATGATAAAGAAAAAGCAGAACGAAAAAGA 960

QY 3459 TTACTAGAAGAGAACACACAGCTTCARAAAGCTCCTGAAGAGAGTAACATTTGACATTT 3518  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 661 TTACTAGAAGAGAGAACACAGCTTCARAAAGCTCCTGAAGAGAGTAACATTTGACATTT 720  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 3519 TATCATAGAGCTTCAAGATTGGGAAGTTCAGCCATGAATATACAAATGCTTTTGGAAAAA 3578  
Db 721 TATCATAGAGCTTCAAGATTGGGAAGTTCAGCCATGAATATACAAATGCTTTTGGAAAAA 780  
QY 3579 CTCAAAGGAATCCCCAGTTAGAAAGGCCAGAGACATGATAACGTGGATCTTCAGATG 3638  
Db 781 CTCAAAGGAATCCCCAGTTAGAAAGGCCAGAGACATGATAACGTGGATCTTCAGATG 840  
QY 3639 TTTGACACAGTGAACCGATTAAAGAGAAAAATCTGTTAAATGTAGCAACACATCAGGA 3698  
Db 841 TTTGACACAGTGAACCGATTAAAGAGAAAAATCTGTTAAATGTAGCAACACATCAGGA 900  
QY 3699 TCGGAATCTATTAAAGATGATGAGATTACTCATGATAAGAAAAAGCAGAAAGAAAGA 3758  
Db 901 TCGGAATCTATTAAAGATGATGAGATTACTCATGATAAGAAAAAGCAGAAAGAAAGA 960  
QY 3759 AAAGCTGAAGCTGTAGGCTACATCGCCAGAGATCATGGC 3799  
Db 961 AAAGCTGAAGCTGTAGGCTTCATCGCCAGAGATCATGGC 1001  
  
RESULT 14  
AF061556  
LOCUS Homo sapiens ubiquitin-protein ligase E3-alpha (UBR1) mRNA, partial cds.  
DEFINITION Homo sapiens ubiquitin-protein ligase E3-alpha (UBR1) mRNA, partial cds.  
ACCESSION AF061556  
VERSION AF061556.1 GI:3170888  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE Kwon,Y.T., Reiss,Y., Fried,V.A., Hershko,A., Yoon,J.K., Gonda,D.K., Sangan,P., Copeland,N.G., Jenkins,N.A. and Varshavsky,A.  
AUTHORS The mouse and human genes encoding the recognition component of the N-end rule pathway  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 95 (14), 7898-7903 (1998)  
MEDLINE 98318583  
PUBMED 9653112  
REFERENCE Kwon,Y.T. and Varshavsky,A.  
AUTHORS Direct Submission  
TITLE Submitted (24-APR-1998) Division of Biology, 147-75, California Institute of Technology, 1200 E. California Bl., Pasadena, CA 91125, USA  
FEATURES  
source Location/Qualifiers  
1..999 /organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/chromosome="15"  
/map="15q15-q21.1"  
1..>999 /gene="UBR1"  
1..>999 /gene="UBR1"  
/note="UBR1p; recognition component of the N-end rule pathway"  
/codon\_start=1  
/product="ubiquitin-protein ligase E3-alpha"  
/protein\_id="AAC39845.1"  
/db\_xref="GI:3170889"  
/translation="MDPNKFLLLVLRVELAEAFNKTISTKDDLIKOYNTLIBEMLQ  
VLIVIGERYVPGVGNVTKREVIREIITLLICIEPMPHSAIAKNIPENNNTGLENV  
INKVATEKPGVSGHGVYELKDESLKDFNMFYHYSKTQHSKAEHMQKRRKQENKDE  
ALPPPPPEFCFAPSFKVINLLNCDIMMYILRTVFERAINTSLNLTGMLQMAFHILA

LCLLBKOOLOKAPBEEVTFDFYHKASRLGSSAMNTOMLLEKLGIPOLBGOXDMITW ILOMEDTVKRLREKSLIVATTTSGSISINDEITHDKERARERKAEARLRHOKIM"	BASE COUNT ORIGIN	363 a 185 c 204 g 247 t	Query Match Best Local Similarity 15.8%; Score 994.2; DB 9; Length 999; Matches 996; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 2799 ATGGATCCCAATAAGTTCTTTTACTTGGTACTTTCAGAGGTATGAACCTTGCAGGCTTTT 2858	QY		
Db 1 ATGGATCCCAACAAGTTCTTTTACTTGGTACTTTCAGAGGTATGAACCTTGCAGGCTTTT 60	Db		
QY 2859 AACAGACCATATCTACAAAAGACCAGGATTTCATTAAACAATATATATACACTAATAGAA 2918	QY		
Db 61 AACAGACCATATCTACAAAAGACCAGGATTTCATTAAACAATATATATACACTAATAGAA 120	Db		
QY 2919 GAAATGCTTCAGGTCCTCATCTATATTTGGGGTGGAGGTATGTACCTTGAGTGGGAAAT 2978	QY		
Db 121 GAAATGCTTCAGGTCCTCATCTATATTTGGGGTGGAGGTATGTACCTTGAGTGGGAAAT 180	Db		
QY 2979 GTGACCAAAAGAGAGGTACAAATGAGAGAAATCATTACCTTGTCTTTCATTGAACCCATG 3038	QY		
Db 181 GTGACCAAAAGAGAGGTACAAATGAGAGAAATCATTACCTTGTCTTTCATTGAACCCATG 240	Db		
QY 3039 CCACACAGTGCCATTGCCAAAAATTTACCTTGAGAATGAAATATGAAATGCTTGCTTAGAG 3098	QY		
Db 241 CCACACAGTGCCATTGCCAAAAATTTACCTTGAGAATGAAATATGAAATGCTTGCTTAGAG 300	Db		
QY 3099 AATGTCATAAACAAGTGGCCACATTTAAGAAACCAGGTGTATCAGGCCATGAGATTAT 3158	QY		
Db 301 AATGTCATAAACAAGTGGCCACATTTAAGAAACCAGGTGTATCAGGCCATGAGATTAT 360	Db		
QY 3159 GAACTAAAAGATGAATCACTTGAAGACTTCAATATGTACTTTTATCATTTACTCTCAAAACC 3218	QY		
Db 361 GAACTAAAAGATGAATCACTTGAAGACTTCAATATGTACTTTTATCATTTACTCTCAAAACC 420	Db		
QY 3219 CAGCATAGCAAGGCTGAACATATGCAGAGAAAAAGGAGAAAAACAAGAAAAACAAGATGAA 3278	QY		
Db 421 CAGCATAGCAAGGCTGAACATATGCAGAGAAAAAGGAGAAAAACAAGAAAAACAAGATGAA 480	Db		
QY 3279 GCATTGCCGCCACCACCTCCTCGAATTCCTGCCCTGCTTTCAGCAAAAGTATTAACTTT 3338	QY		
Db 481 GCATTGCCGCCACCACCTCCTCGAATTCCTGCCCTGCTTTCAGCAAAAGTATTAACTTT 540	Db		
QY 3339 CTCACACTGTGATATCATGTATACATTTCTCAGAGCCGTATTTAGCGCGGAATAGACACA 3398	QY		
Db 541 CTCACACTGTGATATCATGTATACATTTCTCAGAGCCGTATTTAGCGCGGAATAGACACA 600	Db		
QY 3399 GATTCTTAACCTTGTGGACCGAAGGATGCTCCAAATGGCTTTTCATATTTCTGGCATTTGGT 3458	QY		
Db 601 GATTCTTAACCTTGTGGACCGAAGGATGCTCCAAATGGCTTTTCATATTTCTGGCATTTGGT 660	Db		
QY 3459 TTACTAGAAGAGAACCAACAGCTTCAAAAAGCTCTCGAAGAGCAATGATTAACATTTGACTTT 3518	QY		
Db 661 TTACTAGAAGAGAACCAACAGCTTCAAAAAGCTCTCGAAGAGCAATGATTAACATTTGACTTT 720	Db		
QY 3519 TATCATAAAGGCTTCAAGATTGGGAAGTTTCAGCCATGAATATACAAATGCTTTTGGAAAAA 3578	QY		
Db 721 TATCATAAAGGCTTCAAGATTGGGAAGTTTCAGCCATGAATATACAAATGCTTTTGGAAAAA 780	Db		
QY 3579 CTCAAAAGGAATCCCCAGTTAGAAAGGCCAGAGACATGATAAGCTTGATCTTTCAGATG 3638	QY		
Db 781 CTCAAAAGGAATCCCCAGTTAGAAAGGCCAGAGACATGATAAGCTTGATCTTTCAGATG 840	Db		
QY 3639 TTTGACACAGTGAAGCGATTAAAGAAAAATCTTTTAAATTTAGCAACACATCAGGA 3698	QY		
Db 841 TTTGACACAGTGAAGCGATTAAAGAAAAATCTTTTAAATTTAGCAACACATCAGGA 900	Db		
QY 3699 TCGGAATCTATTAAAGATGATGAGATTACTCATGATAAGAAAAAGCAGAAAGAAAGA 3758	QY		
Db 901 TCGGAATCTATTAAAGATGATGAGATTACTCATGATAAGAAAAAGCAGAAAGAAAGA 960	Db		







GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 26, 2003, 19:32:57 : Search time 993.895 Seconds  
(without alignments)  
17132.652 Million cell updates/sec

Title: US-09-724-126A-1

Perfect score: 6308

Sequence: 1 gccagaattcgccagcagg.....aattttgtatttggtgtttt 6308

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_19Jun03.\*

1: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*  
25: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
1	5592	88.6	7742	25	ABZ24689		Human cell growth,
2	4435.4	70.3	6395	20	AAC86933		CDNA encoding mous
3	4435.4	70.3	6395	22	AAC86933		Nucleotide sequenc
4	2546.8	40.4	2550	22	AAH14878		Human CDNA sequenc
5	1311	20.8	1635	24	ABQ75898		Human ubiquitin re
6	1099.4	17.4	6840	22	AAK51709		Human polynucleoti
7	1090	17.3	6850	22	AAK52693		Human polynucleoti
8	996.2	15.8	1001	20	AAK03300		Partial CDNA encod

9	996.2	15.8	1001	22	AAC86934		Nucleotide sequenc
10	813.6	12.9	3327	20	AAX35730		CDNA encoding a pr
11	813.6	12.9	3502	20	AAX35731		Human colon cancer
12	656.8	10.4	733	21	AAH02411		Human CDNA clone (
13	647	10.3	712	22	AAH07621		Human colon cancer
14	638.6	10.1	756	21	AAH02327		Fibrinogen 9,57 co
15	626.8	9.9	2616	24	ABZ70161		DNA encoding novel
16	450.8	7.1	777	23	AAS70068		DNA encoding novel
17	429	6.8	1464	23	AAS70104		Human polynucleoti
18	374.8	5.9	1772	24	ABL90577		DNA encoding novel
19	271.4	4.3	1172	23	AAS70851		Human CDNA encodin
20	250.6	4.0	972	24	ABK35394		Gastric cancer ass
21	246.8	3.9	807	20	AAX39891		Human colon cancer
22	241.6	3.8	264	21	AAA00233		cDNA upregulated i
23	219.8	3.5	4573	25	ACA03886		Drosophila melanog
24	216.8	3.4	6278	23	ABL14799		Gastric cancer ass
25	191.6	3.0	800	20	AAX39892		EST clone DA490.
26	159	2.5	455	20	AAV89674		Human polynucleoti
27	142	2.3	643	24	ABL89546		Drosophila melanog
28	132	2.1	13134	23	ABL14798		Human immune/haema
29	128	2.0	529	22	AAK62305		Human gene signatu
30	116.6	1.8	308	16	AAT21876		Human secreted pro
31	105.6	1.7	365	21	AAC00938		Human secreted pro
32	105.6	1.7	456	21	AAC00940		Human ovarian anti
33	105.6	1.7	1459	24	ABQ54866		Human cancer assoc
34	105.6	1.7	1461	21	AAC78096		Human prostate exp
35	104.8	1.7	507	23	ABV59104		Human novel polynu
36	104.8	1.7	1401	25	ABX05003		Human polynucleoti
37	103.2	1.6	428	22	AAI91546		Human secreted pro
38	103.2	1.6	605	21	AAC00939		Human secreted pro
39	100.8	1.6	602	21	AAC10733		Human secreted pro
40	100.8	1.6	692	21	AAC10734		Human CDNA for nov
41	100.8	1.6	697	24	ABK34403		Human secreted pro
42	100.8	1.6	842	21	AAC10872		Human CDNA differe
43	100.8	1.6	1370	24	ABK83603		Human lung specifi
44	100.6	1.6	1382	25	ABQ82975		Bovine EST associa
45	89.6	1.4	326	25	ABX49787		

#### ALIGNMENTS

RESULT 1  
ABZ24689  
ID ABZ24689 standard; CDNA; 7742 BP.  
XX AC ABZ24689;  
XX DT 07-APR-2003 (first entry)  
XX DE Human cell growth, differentiation and death protein CGDD-1 CDNA.  
XX KW CGDD-1; cell growth; cell differentiation; cell death; human;  
KW cytosolic; antiarteriosclerotic; hepatotropic; antiinflammatory;  
KW antipsoriatic; antianaemic; ophthalmological; auditory;  
KW anticonvulsant; cerebroprotective; nootropic; neuroprotective;  
KW antiparkinsonian; neuroleptic; tranquilizer; immunosuppressive;  
KW anti-HIV; antiallergic; antiasthmatic; antithyroid; antidiabetic;  
KW dermatological; nephrotropic; antirheumatic; antiarthritis;  
KW antitumor; vulnery; virucide; antibacterial; fungicide;  
KW antiparasitic; protozoacide; antihelminthic; antiinfertility;  
KW gynaecological; ubiquitin protein ligase; enzyme; gene therapy;  
KW microarray; gene; ss.  
XX OS Homo sapiens.  
XX FH key Location/Qualifiers  
XX FT CDS 13..5229  
XX FT /\*tag= a  
XX FT /product= "Human CGDD-1"  
XX PN WO200297032-A2.



PD 05-DEC-2002.

XX 05-APR-2002; 2002W0-US11152.

PF 05-APR-2001; 2001US-282110P.

XX 11-APR-2001; 2001US-283294P.

PR 26-APR-2001; 2001US-286820P.

PR 27-APR-2001; 2001US-287228P.

PR 16-MAY-2001; 2001US-291662P.

PR 18-MAY-2001; 2001US-291846P.

PR 25-MAY-2001; 2001US-293727P.

PR 01-JUN-2001; 2001US-295263P.

PR 01-JUN-2001; 2001US-295340P.

PR 15-JAN-2002; 2002US-349705P.

XX (INCYTE GENOMICS INC.

XX Azimzai Y, Au-Young JK, Batra S, Baughn MR, Becha SD, Borowsky ML;

PI Burford N, Ding L, Elliott VS, Emerling BM, Gandhi AR, Gietzen KJ;

PI Griffin JA, Hafalia AJA, Honchell CD, Lal PG, Lee SY, Lu DAM;

PI Arvizu CS, Ramkumar J, Reddy R, Sanjanwala MM, Tang YT, Wallia NK;

PI Wang YE, Warren BA, Xu Y, Yang J, Yao MG, Yue H, Zebbarjadian Y;

XX WPI; 2003-140453/13.

DR P-PSDB; ABP58330.

XX Novel human proteins associated with cell growth, differentiation and

XX death, useful for treating, diagnosing or preventing cancer,

PT developmental, neurological, reproductive or autoimmune/inflammatory

PT disorders -

XX Claim 5; Page 216-218; 238pp; English.

XX The present sequence is that of Incyte clone 1351608CB1 encoding

CC human CGDD-1, a novel protein associated with cell growth, for the

CC differentiation and death. A representative cDNA library for the

CC polynucleotide is PGANN0701 from paragonomic tumour tissue.

CC Structural features establish the encoded protein as being

CC associated with cell growth, differentiation and death, with

CC further evidence suggesting it to be a ubiquitin protein ligase.

CC The invention is based on novel human CGDD-1 to -21 proteins (see

CC ABP58330-507), the polynucleotides encoding them (see AB224689-709),

CC and to the use of these for the diagnosis, treatment or prevention

CC of cell proliferative disorders including cancer, developmental

CC disorders, neurological disorders, autoimmune disorders,

CC reproductive disorders, and disorders of the placenta, and in the

CC assessment of the effects of exogenous compounds on the activity

CC and expression of proteins and nucleic acids associated with cell

CC growth, differentiation and death. CGDD polynucleotides are also

CC used in a claimed microarray and in a claimed method of generating

CC an expression profile of a sample.

XX Sequence 7742 BP; 2388 A; 1423 G; 1622 G; 2309 T; 0 other;

XX

Query Match 88.6%; Score 5592; DB 25; Length 7742;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 5592; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 717 GGTAAGTACAGAGGATGAAATACAGCGGGAGTATACCCAGACCCCTACAGCGTCTGGCATCT 776

DB 1 GGTAAGTACAGAGGATGAAATACAGCGGGAGTATACCCAGACCCCTACAGCGTCTGGCATCT 60

QY 777 TGGTGGGATCAGCAAGTGGATTTTATACATGCTTTCTTGGCATCATTTGGCACAATTTGGT 836

DB 61 TGGTGGGATCAGCAAGTGGATTTTATACATGCTTTCTTGGCATCATTTGGCACAATTTGGT 120

QY 837 CCAGAAATTTTACATTTTGTGTAATGGGACCCAGACTTGGAAAAGCAGGAGGAAAGTGTACAA 896

DB 121 CCAGAAATTTTACATTTTGTGTAATGGGACCCAGACTTGGAAAAGCAGGAGGAAAGTGTACAA 180

QY 897 ATGTCAATATTCACCTCCAGTGAATGGTACTTATTTGGAGAAGATCCAGATATTTGGCTTA 956

DB 181 ATGTCAATATTCACCTCCAGTGAATGGTACTTATTTGGAGAAGATCCAGATATTTGGCTTA 240

QY 957 GAGAAATTGAAGACAGGTGGAGCATTTTCAGCTTTTGTGGAGGGTTTTCAAAAGTGGAGAG 1016

DB 241 GAGAAATTGAAGACAGGTGGAGCATTTTCAGCTTTTGTGGAGGGTTTTCAAAAGTGGAGAG 300

QY 1017 ACAACCTATTCTTGCAGGGATTGCAATTTGATCAACATGTGTACTCTGTATGACTGCG 1076

DB 301 ACAACCTATTCTTGCAGGGATTGCAATTTGATCAACATGTGTACTCTGTATGACTGCG 360

QY 1077 TTCCAGGACAGTGTTCATAAAAAATCATCGTTACAAGATGCATACCTTCTACTCGAGGAGG 1136

DB 361 TTCCAGGACAGTGTTCATAAAAAATCATCGTTACAAGATGCATACCTTCTACTCGAGGAGG 420

QY 1137 TTCTGTGACTGTGGAGACACAGAGGCATGGAAAACTGGCCCCCTTTTGTGTAAATCATGAA 1196

DB 421 TTCTGTGACTGTGGAGACACAGAGGCATGGAAAACTGGCCCCCTTTTGTGTAAATCATGAA 480

QY 1197 CCTGGAAGACAGGTACTATTAAGAGAAATTCACGCTGTCCGTTGTAAGTGAAGAGTAATT 1256

DB 481 CCTGGAAGACAGGTACTATTAAGAGAAATTCACGCTGTCCGTTGTAAGTGAAGAGTAATT 540

QY 1257 GTCCAAGCCAGAAAAATTTTCCCTTCAGTGATAAAATATGCTGTAAGAAATGACTATATGG 1316

DB 541 GTCCAAGCCAGAAAAATTTTCCCTTCAGTGATAAAATATGCTGTAAGAAATGACTATATGG 600

QY 1317 GAAGAGAAAAAAGAACTGCCCTCCGAACTCCAGATAAGGGAGAAAAATGAAAGATACTAT 1376

DB 601 GAAGAGAAAAAAGAACTGCCCTCCGAACTCCAGATAAGGGAGAAAAATGAAAGATACTAT 660

QY 1377 TGTGTCTCTTTCAATGATGAACACCATTTTCATATGACACGCTCATATACAGCCTCAAGA 1436

DB 661 TGTGTCTCTTTCAATGATGAACACCATTTTCATATGACACGCTCATATACAGCCTCAAGA 720

QY 1437 GCTCTTGACTGTGAGCTCGCAGAGCCAGTTCATACCATCCACTGCCATTCACAAAGAGGT 1496

DB 721 GCTCTTGACTGTGAGCTCGCAGAGCCAGTTCATACCATCCACTGCCATTCACAAAGAGGT 780

QY 1497 CGTCGGGGTGTAAAGCGGGAGCTTTATGCTGCTGCCAGGAAGCAAGAAAGATATAAG 1556

DB 781 CGTCGGGGTGTAAAGCGGGAGCTTTATGCTGCTGCCAGGAAGCAAGAAAGATATAAG 840

QY 1557 AGTCATTCAGAAAAATGCTCTCAACATCCATTCATGATAGAGTATTCACATCAGAGATT 1616

DB 841 AGTCATTCAGAAAAATGCTCTCAACATCCATTCATGATAGAGTATTCACATCAGAGATT 900

QY 1617 ATGCTCATCAGAAATTTGCTTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1676

DB 901 ATGCTCATCAGAAATTTGCTTTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960

QY 1677 TCAAGTGAATTTAGGCAGATCTTTTCCCAAGCATGCTTTAGAGAGAAAGCTGACCTCGGAG 1736

DB 961 TCAAGTGAATTTAGGCAGATCTTTTCCCAAGCATGCTTTAGAGAGAAAGCTGACCTCGGAG 1020

QY 1737 AATCCCTGCTCATTAAGCAGGTTAATGCTTTGGGATGCAAGCTTTTAAAGGTGCCCT 1796

DB 1021 AATCCCTGCTCATTAAGCAGGTTAATGCTTTGGGATGCAAGCTTTTAAAGGTGCCCT 1080

QY 1797 AAGATCCTTTCATGAATTTGCTTTCAGCAGTTTTATGAGAGATGGAATACAAAACATC 1856

DB 1081 AAGATCCTTTCATGAATTTGCTTTCAGCAGTTTTATGAGAGATGGAATACAAAACATC 1140

QY 1857 TTTGCTATGGAATTTGTGAAGTATTTAAACAACTGCAAGAAATATATCATGATGATAT 1916

DB 1141 TTTGCTATGGAATTTGTGAAGTATTTAAACAACTGCAAGAAATATATCATGATGATAT 1200

QY 1917 CATGACAGAGTATCTCTATAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1976

DB 1201 CATGACAGAGTATCTCTATAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1260

QY 1977 GCTGACATCTTATGAGAGCAGAAATGTTATCTCTGTCATTTACTGAAACTCTGCTAGAA 2036

DB 1261 GCTGACATCTTATGAGAGCAGAAATGTTATCTCTGTCATTTACTGAAACTCTGCTAGAA 1320

2037 GTTTTACCTGAGTACTTGGACAGGAACAATAAATTCACCTTCCAGGGTTATAGCCAGGAC 2096  
1321 GTTTTACCTGAGTACTTGGACAGGAACAATAAATTCACCTTCCAGGGTTATAGCCAGGAC 1380  
2097 AAATTTGGGAAGATATATGACGTAATATGACCTAAAGATATATCCTGATCAGCAAAACC 2156  
1381 AAATTTGGGAAGATATATGACGTAATATGACCTAAAGATATATCCTGATCAGCAAAACC 1440  
2157 ACAATATGGACAGAAATTAAGAAATCAGTTCCTTGAAGGTTTTCGATCTTTTGAAG 2216  
1441 ACAATATGGACAGAAATTAAGAAATCAGTTCCTTGAAGGTTTTCGATCTTTTGAAG 1500  
2217 ATTCTTACCTGTATGACAGGAATGGAAGAAATCGAAGACAGGTTGGCAACACATTTGAA 2276  
1501 ATTCTTACCTGTATGACAGGAATGGAAGAAATCGAAGACAGGTTGGCAACACATTTGAA 1560  
2277 GTGGATCTGATTTGGGAGGCTGCCATTTGATGAAGAACTCTTACTTGTGGCTTATAAAGATGT 2336  
1561 GTGGATCTGATTTGGGAGGCTGCCATTTGATGAAGAACTCTTACTTGTGGCTTATAAAGATGT 1620  
2337 ATGTTTCAAGAGTGGTGTGCTTGTGATGAAGAACTCTTACTTGTGGCTTATAAAGATGT 2396  
1621 ATGTTTCAAGAGTGGTGTGCTTGTGATGAAGAACTCTTACTTGTGGCTTATAAAGATGT 1680  
2397 CACAAAGCTGTGATGAGTGCAGTACCAGTTTCATATCTAGTACCAAGACAGTAGTACAA 2456  
1681 CACAAAGCTGTGATGAGTGCAGTACCAGTTTCATATCTAGTACCAAGACAGTAGTACAA 1740  
2457 TCGTGTGGACATAGTTTGGAAACAAAGTCCACAGAGTATCTGAGGATCTTGTAAAGATA 2516  
1741 TCGTGTGGACATAGTTTGGAAACAAAGTCCACAGAGTATCTGAGGATCTTGTAAAGATA 1800  
2517 CATCTGCCACTCTCTAGGACCCCTTGTGCTTCAATGCTTCAAGCAGGCTGGGTGCT 2576  
1801 CATCTGCCACTCTCTAGGACCCCTTGTGCTTCAATGCTTCAAGCAGGCTGGGTGCT 1860  
2577 GTTTCAGACTGATGAATTTGCTTTTGGAGACTTTCAGAGTAGAGTACTAGTGGAA 2636  
1861 GTTTCAGACTGATGAATTTGCTTTTGGAGACTTTCAGAGTAGAGTACTAGTGGAA 1920  
2637 TATCCCTTACCTTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2696  
1921 TATCCCTTACCTTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1980  
2697 GGACTGCTCTTATAGCCAGGCTGCTTATACCAAGATGTTAAGTCAGAGAGAAATG 2756  
1981 GGACTGCTCTTATAGCCAGGCTGCTTATACCAAGATGTTAAGTCAGAGAGAAATG 2040  
2757 TATGATAAAGATATCATGCTTTCAGATTTGGTGCATCTTTAATGGATCCCAATAGTTC 2816  
2041 TATGATAAAGATATCATGCTTTCAGATTTGGTGCATCTTTAATGGATCCCAATAGTTC 2100  
2817 TTGCTTACTGCTTACAGAGGTATGACTTGGCGAGGCTTTTACAGAGCATATCTACA 2876  
2101 TTGCTTACTGCTTACAGAGGTATGACTTGGCGAGGCTTTTACAGAGCATATCTACA 2160  
2877 AAAGACAGGATTTGATTAACCAATATAATACACTAATAGAGAAATGCTTCAGGTCCTC 2936  
2161 AAAGACAGGATTTGATTAACCAATATAATACACTAATAGAGAAATGCTTCAGGTCCTC 2220  
2937 ATCTATATTTGGGTGAGCCTTATGTACCTGGAGTGGAAATGTGACCAAGAGAGGTC 2996  
2221 ATCTATATTTGGGTGAGCCTTATGTACCTGGAGTGGAAATGTGACCAAGAGAGGTC 2280  
2997 ACAATGAGAGAAATCATCTACTTGTGATTAACCAATATAATACACTAATAGAGAAATG 3056  
2281 ACAATGAGAGAAATCATCTACTTGTGATTAACCAATATAATACACTAATAGAGAAATG 2340  
3057 AAAAAATTTACCTGAGATGAATTAATGAAGTGGCTTACAGATGTCATTAACCAAGAGT 3116  
2341 AAAAAATTTACCTGAGATGAATTAATGAAGTGGCTTACAGATGTCATTAACCAAGAGT 2400  
3117 GCCACATTTAAGAAACAGGCTGTATCAGGCCATGGAGTTTATGAACATAAAGATGAATCA 3176

2401 GCCACATTTAAGAAACAGGCTGTATCAGGCCATGGAGTTTATGAACATAAAGATGAATCA 2460  
3177 CTGAAGACTTCAATATATGTACTTTTATCATTTACTCCAAAACCCAGCATAGCAAGGTGAA 3236  
2461 CTGAAGACTTCAATATATGTACTTTTATCATTTACTCCAAAACCCAGCATAGCAAGGTGAA 2520  
3237 CATATGCAGAGAAAGGAGAAACAAGAAACAAGATGAAGCATTCGCCACCAACCA 3296  
2521 CATATGCAGAGAAAGGAGAAACAAGAAACAAGATGAAGCATTCGCCACCAACCA 2580  
3297 CCTCCTGAATTCGCTCCTTTCAGCAAGGATTAACCTTCTCAACTGTGTATATCATG 3356  
2581 CCTCCTGAATTCGCTCCTTTCAGCAAGGATTAACCTTCTCAACTGTGTATATCATG 2640  
3357 ATGTACATTTCTCAGGACCGTATTTGAGCGGCAATAGACACAGATTTCAACTTGTGGACC 3416  
2641 ATGTACATTTCTCAGGACCGTATTTGAGCGGCAATAGACACAGATTTCAACTTGTGGACC 2700  
3417 GAAGGATGCTCCAAATGGCTTTTATATTTCTGGCATTTGGTTTACTAGAGAGAGCA 3476  
2701 GAAGGATGCTCCAAATGGCTTTTATATTTCTGGCATTTGGTTTACTAGAGAGAGCA 2760  
3477 CAGCTTCAAAAGCTCCTGAAGAAAGTAAATTTGACTTTTATCATAGGCTTCAAGA 3536  
2761 CAGCTTCAAAAGCTCCTGAAGAAAGTAAATTTGACTTTTATCATAGGCTTCAAGA 2820  
3537 TTGGAGGCTCAGCCATGAATATACAAATGCTTTTGGAAAACTCAAGGAATTCGCCAG 3596  
2821 TTGGAGGCTCAGCCATGAATATACAAATGCTTTTGGAAAACTCAAGGAATTCGCCAG 2880  
3597 TTGAAGGCCAGAGACATGATACGTTGATCTTACAGATTTGACACAGTGAAGCGA 3656  
2881 TTGAAGGCCAGAGACATGATACGTTGATCTTACAGATTTGACACAGTGAAGCGA 2940  
3657 TTAAGAGAAATCTTGTAAATTTAGCAACACATCAGGATCGGAATCTATTAGAAT 3716  
2941 TTAAGAGAAATCTTGTAAATTTAGCAACACATCAGGATCGGAATCTATTAGAAT 3000  
3717 GATGAGATTACTCATGATAAAGAAAGCAGAAAGAAAGCTGAAGCTGCTAGG 3776  
3001 GATGAGATTACTCATGATAAAGAAAGCAGAAAGAAAGCTGAAGCTGCTAGG 3060  
3777 CTACATGCCAGAGATCATGCTCAGATGTCTGCTTACAGAAAACTTCAATTGAACT 3836  
3061 CTACATGCCAGAGATCATGCTCAGATGTCTGCTTACAGAAAACTTCAATTGAACT 3120  
3837 CATAACTCATGTATGACAAATACATCAGAAATGCTTGGAAAGAAATTCATTTGGAG 3896  
3121 CATAACTCATGTATGACAAATACATCAGAAATGCTTGGAAAGAAATTCATTTGGAG 3180  
3897 GAAGAGACCCCGCAGCTGACTACTCTAGAAATGCTTTGGGTCTCTAAACGGGT 3956  
3181 GAAGAGACCCCGCAGCTGACTACTCTAGAAATGCTTTGGGTCTCTAAACGGGT 3240  
3957 CCATCTGTACTGAAAGAGGCTGCTGAGTGCATCTTGGCAAGAAAGAGGAGTG 4016  
3241 CCATCTGTACTGAAAGAGGCTGCTGAGTGCATCTTGGCAAGAAAGAGGAGTG 3300  
4017 AAAATAGAAAAATAATGCCATGTTATCGGGCTGTGTCCAGAAATCTACTGCTTTAACC 4076  
3301 AAAATAGAAAAATAATGCCATGTTATCGGGCTGTGTCCAGAAATCTACTGCTTTAACC 3360  
4077 CAGCACAGGGGAAACCCATAGAACTCTCAGAGAGGCTTAGACCCACTTTTCATGGAT 4136  
3361 CAGCACAGGGGAAACCCATAGAACTCTCAGAGAGGCTTAGACCCACTTTTCATGGAT 3420  
4137 CCAGACTTGGCATATGGAACTTATACAGGAAGCTGTGGTCATGTATGACCGCAGTGTGC 4196  
3421 CCAGACTTGGCATATGGAACTTATACAGGAAGCTGTGGTCATGTATGACCGCAGTGTGC 3480  
4197 TGGCAGAAAGTATTTGAAGCTGTACAGCTGAGCTCTCAGCAGCGCATTCATGTTGACCTT 4256

Db	3481	TGCGAAGATATTTTGAAGCTGTACAGCTGAGCTCTCAGACGCGCATTTTCATGTTGACCTT	3540
Qy	4257	TTTGACATTGGAAGTGGAGATATCTTTGGCCCTCTTGCAAAATCTCTGTGCAATACACTGCTG	4316
Db	3541	TTTGACATTGGAAGTGGAGATATCTTTGGCCCTCTTGCAAAATCTCTGTGCAATACACTGCTG	3600
Qy	4317	ATCCCCATTTATCCCTTTGCAACCTCAAAAGATAACAGTACAGTACAGATGCGCTTTGCT	4376
Db	3601	ATCCCCATTTATCCCTTTGCAACCTCAAAAGATAACAGTACAGTACAGATGCGCTTTGCT	3660
Qy	4377	CAACTTTTGACCTTGGCACCGTGGATACAGACTCTTCTGGCCAGAAATATCAGGTATATAAT	4436
Db	3661	CAACTTTTGACCTTGGCACCGTGGATACAGACTCTTCTGGCCAGAAATATCAGGTATATAAT	3720
Qy	4437	ATAAGACATGCTAAAGGAGAAAACCCAAATTCCTATTTTCTTTAATCAAGGAATGGAGAT	4496
Db	3721	ATAAGACATGCTAAAGGAGAAAACCCAAATTCCTATTTTCTTTAATCAAGGAATGGAGAT	3780
Qy	4497	TCTACTTTGGAGTTCCATTCCTCAGTTTGGCGTTGAGTCTTCGATTTAAATATCA	4556
Db	3781	TCTACTTTGGAGTTCCATTCCTCAGTTTGGCGTTGAGTCTTCGATTTAAATATCA	3840
Qy	4557	AATAGCATCAAGGAATGGTATTTCTTTGCCACAAATTTATAGAAATTTGGATTTGAAA	4616
Db	3841	AATAGCATCAAGGAATGGTATTTCTTTGCCACAAATTTATAGAAATTTGGATTTGAAA	3900
Qy	4617	GTGCCACTGATGAAGGATTCCTCGAGTCCCTATGCTGACCTGGAGCACTCGCGCTTTC	4676
Db	3901	GTGCCACTGATGAAGGATTCCTCGAGTCCCTATGCTGACCTGGAGCACTCGCGCTTTC	3960
Qy	4677	ACTATCAGGCAATTTGAAATCTATTTGGGAGATCAAGGAAACCTCTGTTTGGAGCACTT	4736
Db	3961	ACTATCAGGCAATTTGAAATCTATTTGGGAGATCAAGGAAACCTCTGTTTGGAGCACTT	4020
Qy	4737	CAAAATAGGCAATATGCTGCTCAAGACATTAATGCAAGTTTGAGTTGCACAGAGGATT	4796
Db	4021	CAAAATAGGCAATATGCTGCTCAAGACATTAATGCAAGTTTGAGTTGCACAGAGGATT	4080
Qy	4797	ACCTGTCTCAGGCTCTGATACAGAAACATCTGCTTCTCTATCAGTTTGTCTTCTCT	4856
Db	4081	ACCTGTCTCAGGCTCTGATACAGAAACATCTGCTTCTCTATCAGTTTGTCTTCTCT	4140
Qy	4857	AACATATAATCAGAAATACACCATGCTTCTGCTATAGATCTGTTTCATGTTTGGTG	4916
Db	4141	AACATATAATCAGAAATACACCATGCTTCTGCTATAGATCTGTTTCATGTTTGGTG	4200
Qy	4917	GGTGTGTGTAGCATTTCCCATCTCTGTTGGGATGACCTGTTGATCTCGACCTTCT	4976
Db	4201	GGTGTGTGTAGCATTTCCCATCTCTGTTGGGATGACCTGTTGATCTCGACCTTCT	4260
Qy	4977	TCAGTTAGTTCTTCTATACCACTTTATCTCTTCCATTTGATCACCATGGCACACATG	5036
Db	4261	TCAGTTAGTTCTTCTATACCACTTTATCTCTTCCATTTGATCACCATGGCACACATG	4320
Qy	5037	CTTCAGATACACTACAGTACAGACAGGCTTACCCCTTGTCTGAGTTTCAAGAGACAGT	5096
Db	4321	CTTCAGATACACTACAGTACAGACAGGCTTACCCCTTGTCTGAGTTTCAAGAGACAGT	4380
Qy	5097	GAAGGCTCATTCGGCATCTCTTCTTTCAGAGAAATTTCTCAATATACAGTGGCTTCC	5156
Db	4381	GAAGGCTCATTCGGCATCTCTTCTTTCAGAGAAATTTCTCAATATACAGTGGCTTCC	4440
Qy	5157	ATTGGGTGTATATTCCTGGCTGTTATTTGTGGGTCTCACATGAAGAAATGGCATCACCT	5216
Db	4441	ATTGGGTGTATATTCCTGGCTGTTATTTGTGGGTCTCACATGAAGAAATGGCATCACCT	4500
Qy	5217	TATCTTCGCTGTGCTGCTTTTTCACATATTTTACCTTGGGTAACTCCGCTCAGGAA	5276
Db	4501	TATCTTCGCTGTGCTGCTTTTTCACATATTTTACCTTGGGTAACTCCGCTCAGGAA	4560
Qy	5277	CTGCATACCAATTTCTGCAGAGGAGATACAGTGCATCTGTAGCTATCTATCTTTACCT	5336
Db	4561	CTGCATACCAATTTCTGCAGAGGAGATACAGTGCATCTGTAGCTATCTATCTTTACCT	4620

RESULT 2  
AAV99308  
ID AAV99308 standard; cdNA; 6395 BP.

Qy	5337	ACAAATTTTGTCTCTCTCTTCCAGGAATATTTGGGATACTGTAAAGCCCTTGTCTCCAGAGG	5396
Db	4621	ACAAATTTTGTCTCTCTCTTCCAGGAATATTTGGGATACTGTAAAGCCCTTGTCTCCAGAGG	4680
Qy	5397	TGGTGTGCAGATCCCTTCTTACTAACTGTTTGAAGCAAAACACACCTGCTGTCAGGTAC	5456
Db	4681	TGGTGTGCAGATCCCTTCTTACTAACTGTTTGAAGCAAAACACACCTGCTGTCAGGTAC	4740
Qy	5457	CCTAGAAAAGAAATAGTTTGTATAGAGCTTCTCTGATGACTATAGCTGCTCTCTGAATCAA	5516
Db	4741	CCTAGAAAAGAAATAGTTTGTATAGAGCTTCTCTGATGACTATAGCTGCTCTCTGAATCAA	4800
Qy	5517	GCTTCTCATTTTCAGGTGCCACGGTCTGCAGATGATGAGCGAAGCATCCTGTCTCTGTC	5576
Db	4801	GCTTCTCATTTTCAGGTGCCACGGTCTGCAGATGATGAGCGAAGCATCCTGTCTCTGTC	4860
Qy	5577	CTTTTCTGTGGGCTATACTATGTTCTCAGAACATTTGCTGTCAGGAAATTTGTGAACGGG	5636
Db	4861	CTTTTCTGTGGGCTATACTATGTTCTCAGAACATTTGCTGTCAGGAAATTTGTGAACGGG	4920
Qy	5637	GAAGAGTTGGAGCTTGCAATTTTTCACGCACCTTCACTGTGGAGCCGAGTCTGCAATTTTC	5696
Db	4921	GAAGAGTTGGAGCTTGCAATTTTTCACGCACCTTCACTGTGGAGCCGAGTCTGCAATTTTC	4980
Qy	5697	CTAAAAATCAGAGAATCGCGAGTGTCTGTGTTGAAGTTAAAGCCAGAGGCTGTGCCTAT	5756
Db	4981	CTAAAAATCAGAGAATCGCGAGTGTCTGTGTTGAAGTTAAAGCCAGAGGCTGTGCCTAT	5040
Qy	5757	CCAGCTCTCTTACTTGGATGAATATGAGAAACAGACCCCTGGCTGGAAGAGGGCAACCCC	5816
Db	5041	CCAGCTCTCTTACTTGGATGAATATGAGAAACAGACCCCTGGCTGGAAGAGGGCAACCCC	5100
Qy	5817	CTTCAATTTATCTGCTGAGCGGTATCGGAAAGTCCCATTTGCTGTGGCAACAACTGCATTT	5876
Db	5101	CTTCAATTTATCTGCTGAGCGGTATCGGAAAGTCCCATTTGCTGTGGCAACAACTGCATTT	5160
Qy	5877	ATAGAAGAGATTGCTAGGAGCCAAAGACTAATACAGATGTTATTTGGATTCAACTGGCAG	5936
Db	5161	ATAGAAGAGATTGCTAGGAGCCAAAGACTAATACAGATGTTATTTGGATTCAACTGGCAG	5220
Qy	5937	TTACTGTGAGCTTCCAACTCTGCTCAAGACAATCACAATGACGACAGTAGTAAAGGCTG	5996
Db	5221	TTACTGTGAGCTTCCAACTCTGCTCAAGACAATCACAATGACGACAGTAGTAAAGGCTG	5280
Qy	5997	ATTCAAAATTTATGGAAGCTTTCTGAGGCTGGGAAAGTATTTGGAGGCTCTTTTGTCTCCA	6056
Db	5281	ATTCAAAATTTATGGAAGCTTTCTGAGGCTGGGAAAGTATTTGGAGGCTCTTTTGTCTCCA	5340
Qy	6057	TGTCAGGTTCACTTTACATCAATAAATAATTTCTTAATGAGATTTTGTCTTCAATTAGCA	6116
Db	5341	TGTCAGGTTCACTTTACATCAATAAATAATTTCTTAATGAGATTTTGTCTTCAATTAGCA	5400
Qy	6117	AACATATGCTTACAGGAAAAAGGACATAGATCAATCTGTTTATGTCGTAGTATTTCC	6176
Db	5401	AACATATGCTTACAGGAAAAAGGACATAGATCAATCTGTTTATGTCGTAGTATTTCC	5460
Qy	6177	AGGAATTTATTCCTTCAATAATTTGCTCTCATTTTATTTATTTTCATCCACTTTGGTAGAT	6236
Db	5461	AGGAATTTATTCCTTCAATAATTTGCTCTCATTTTATTTATTTTCATCCACTTTGGTAGAT	5520
Qy	6237	GAAGTCACGTCAAAACAGTTTGTAGACATTTTATGTTGGTTAACTCTTCTGCAATTTTGT	6296
Db	5521	GAAGTCACGTCAAAACAGTTTGTAGACATTTTATGTTGGTTAACTCTTCTGCAATTTTGT	6296
Qy	6297	ATTTGGTGTGTTT 6308	
Db	5581	ATTTGGTGTGTTT 5592	

XX AAV99308;  
XX AC  
XX 25-MAR-1999 (first entry)  
XX DE  
XX DNA encoding mouse a ubiquitin-protein ligase, Ubrl.  
XX KW Ubiquitin-protein ligase; Ubrl; mouse; ubiquitinylation; degradation;  
XX KW N-end rule pathway; stress-related muscle wasting; inhibitor; screen; ss.  
XX OS  
XX Mus sp.  
XX FH  
XX Key Location/Qualifiers  
XX FT 115..5388  
XX FT /\*tag= a  
XX FT /product= Ubrl  
XX PN  
XX US5861312-A.  
XX XX  
XX 19-JAN-1999.  
XX XX  
XX 02-DEC-1997; 97US-0982956.  
XX PF  
XX 02-DEC-1997; 97US-0982956.  
XX PR  
XX (CALY ) CALIFORNIA INST OF TECHNOLOGY.  
XX PA  
XX Kwon YT, Varshavsky A;  
XX PI  
XX WPI; 1999-130395/11.  
XX DR P-PSDB; AAW84351.  
XX XX  
XX PT Mouse and human Ubrl cDNA - useful for producing recombinant Ubrl  
XX PT polypeptides  
XX PS  
XX Claim 1; Columns 15-28; 18pp; English.  
XX CC  
XX The present sequence encodes a ubiquitin-protein ligase called Ubrl.  
XX CC The Ubrl enzymes are involved in protein ubiquitinylation and  
XX CC ultimate degradation through the N-end rule pathway and have been  
XX CC linked to stress-related muscle wasting. Recombinant Ubrl polypeptides  
XX CC can be used to screen for inhibitors of muscle wasting when this is  
XX CC associated with the N-end rule pathway.  
XX XX  
XX Sequence 6395 BP; 1802 A; 1376 C; 1519 G; 1698 T; 0 other;  
XX SQ  
Query Match 70.3%; Score 4435.4; DB 20; Length 6395;  
Best Local Similarity 86.3%; Pred. No. 0;  
Matches 4997; Conservative 0; Mismatches 721; Indels 69; Gaps 6;  
QY 590 TTCAGGGCGCGTAAAGTGTGCTGCTGCTC-TCCGACGGGCCACAGGTTCCGCT 648  
DB 8 TTCAGGGCGCGTAAAGTGTGCTGCTGCTGCTGCGCGCGGCCACAGGTTCCGCT 67  
QY 649 TGCCTCTGCGCGGGGTGCGCAACTCGAGGTCAGTTTCCCTCAAGATGGCGGAGG 708  
DB 68 AGCTGCGCGCGGGGTGCGCAACTCGGCGGTCGTTTCCCTTAAAGATGGCGGAGG 127  
QY 709 AGGCTGGAGGTACTGAGAGATGGAATCAGCGCGGAGTTACCCAGACCCCTCAGGTC 768  
DB 128 AGATGACGCCGCCGAGAGATGGAGCTCAGCCCGGAGGCTCCCTGGCCCGCAGCGGC 187  
QY 769 TGGCATCTGGTGGGATCAGCAAGTTGATTTTATATCTGCTTTCATCATTTTGGCAC 828  
DB 188 CGGCATCTGGTGGGATCAGCAAGTTGATTTTATATCTGCTTTCATCATTTTGGCAC 247  
QY 829 AATTGTCGCGAGAAATTTACTTGTCTGAAATGGACCCAGACTTGGAAAAGCAGGAGAA 888  
DB 248 AATTAGTCCGAGAAATTTATTTGCTGAGATGGACCCAGATTTGGAAGCAGAGAGA 307  
QY 889 GTGTCAATGCAATATTCCTCCACTCGAATGGTACTTATTTGGAGAGATCCAGATA 948  
DB 308 GTGTACAGATGTCAATACCTCCTCTTGGAGTGGTACTTATTTGGAGAGATCCGGATA 367

QY 949 TTTGCTTAGAGAAATTTGAAGCACAGTGGAGCATTTTCAGCTTTTGGGAGGTTTTCAAA 1008  
DB TTTGCTTAGAGAAATTTAAACACAGTGGAGCGTTCCAGTGTGTGGGAGGTTTTCAAA 427  
QY 1009 GTGGAGACAACTTATTTTCGAGGGATTGTGCAATTGATCCAACTGTGTACTCTGTA 1068  
DB GTGGAGAAACAATATTTCTGTAGGATTGTGCAATTGATCCAACTGTGTGCTCTGTA 487  
QY 1069 TGGACTGCTCCAGGACAGTGTTCATAAAATCATCGTTACAGATGATCTTACTG 1128  
DB TGGACTGCTCCAAAGTAGTGTTCATAAAACCACATCGTTACAGATGATCTTACTG 547  
QY 1129 GAGGAGGCTGTGTGACTGTGGAGACAGAGGATGGAATACTGGCCCTTTTGTGTA 1188  
DB GAGGGGCTTCTGTGACTGTGGAGACAGAGCGTGGAAAAGTGGCCCTTTTGTGTTG 607  
QY 1189 ATCAATGAACCTTGAAGAGCAGTACTATAAAGAGAATTCACGCTGTCCGTTGAATGA 1248  
DB ATCAGAGCCTTGAAGAGCAGTACTACAAAGAGAGCTTACATTTGCCCATTTGAATGA 667  
QY 1249 AGGTAATGTCCAAAGCCAGGAAATATTTCTTCAGTGAATAATATGTCTGTAAGTA 1308  
DB AGGTATTTGCTCAAGCCAGGAGATATTTCCCTTCGTTGATATAAATACATTTGTA 727  
QY 1309 CTATATGGGAGAGGAAAAGAACTGCTCTGAACTCCAGATAGGAGGAAAATGAA 1368  
DB CTATATGGGAGAGGAAAAGAAATTTGCTCTGAACTCCAGATAGGAGGAAAATGAA 787  
QY 1369 GATATATTTGCTCTTTTCAATGATGAACACCATTTCAATGATGACCATCATATAC 1428  
DB GATATATTTGCTCTTTTCAACGATGAGCACCATTCTGTATGATCATGTGATCTAC 847  
QY 1429 TACAAGAGCTCTTTGACTGTGAGCTCGCAGAGGCCAGTTGATACCATTCGCA 1488  
DB TGCAGAGAGCTCTAGATTGCGAGCTTGCAGAGGCACAGCTGCACAGCTGCATCGA 907  
QY 1489 AAGAGGTGCTGCGGCTGTAAAGGGGAGCTTATGCTGTTCGCTGCGAGGAAAGGA 1548  
DB AAGAGGTGCTGCGGCTGTAAAGCAGGTGTATGCTGCTGCGCTTGGCTTGGATGA 967  
QY 1549 ATATAAGAGTCATTCAGAAAATGCTCTCAACATCCATTCATCTAGAAGTATTAC 1608  
DB ATATAAGAGTCATTCAGAAAATGCTCTCAACATCCATTCATCTAGAAGTATTAC 1027  
QY 1609 CAGAGATTATGGCTCATCAGAAATTTGCTTTCGCTTTCGCTTTCGCTGGAACA 1668  
DB CCGTGGTATGGCTCACCAGAAATTCGCTTTCGCTTTCGCTTTCGCTGGAACA 1087  
QY 1669 TGAGCTATTCAGTGTACTTTAGGAGATCTTTTGCCTCAAGCATGCTTTAGAGA 1728  
DB TGAGCTATTCAGTGTACTTTAGGAGATCTTTTGCCTCAAGCATGCTTTAGAGA 1147  
QY 1729 ACTCGGAGATCCCTGCTCATAGCAGGTTAATGCTTTGGATGCAAGCTTTTAA 1788  
DB GCTCTGAAAATCCCTGCTTATAAGCAGTAAATGCTTTGGGATGCAAACTTTTAA 1207  
QY 1789 GTGCCCTTAAGATCTTCATGAATTTGATCTTTCAGCAGTTTTATGAGATGGA 1848  
DB GTGCCCTTAAGATCTTCATGAATTTGATCTTTCAGCAGTTTTATGAGATGGA 1267  
QY 1849 AAAAATCTTTTGTCTATGGAATTTTGAAGTATTATAAACAATGCGAGAAAATAT 1908  
DB AAAAATCTTTTGTCTATGGAATTTTGAAGTATTATAAACAATGCGAGAAAATAT 1327  
QY 1909 GTGATGATCAGCAGAAATCTCTATAACTGCACTTTTCACTGATGATTTTACTG 1968  
DB GCGACGACCCAGCAGAGAGCATCTCCATAACCGCTCTGCTGCGAGATGCTCAC 1387  
QY 1969 CTACTCTGGCTCGACATCTTATGAAGAGCAGATGTTATCTCTCTCATTTACT 2028  
DB CGACTTTGGCCCGCATCTTATTGAAGAGCAGATGTTATTTCTCTCTCATTTACT 1447

QY 2029 TGCTAGAAGTTTACCTGAGTACTTGGACAGGAACAATAAATTCAACTTCCAGGTTATA 2088  
DB 1448 TGCTAGAAGTTTACCTGAGTACTTGGACAGGAACAATAAATTCAACTTCCAGGTTATA 1507  
QY 2089 GCCAGGACAATGGGAAGAGTATATGACAGTAATATGACCTAAAGTATATCTGATCA 2148  
DB 1508 GCCAGGACAATGGGAAGAGTATATGACAGTAATATGACCTAAAGTATATCTGATTA 1567  
QY 2149 GCAAAACCCACAATATGACAGAAAGATTAAGAATGAGTTCCTTGAAGGTTTTCGATCTT 2208  
DB 1568 GCAAGCCTGTATATGACAGAAAGATTAAGAAGCGCATGTCCTGGAAGGTTTCCGGCTT 1627  
QY 2209 TTTTGAAGATTCTTACCTGTATGACAGGAATGGAAGAAATCCGAAGACAGTTCGGCAAC 2268  
DB 1628 TTTTGAAGATTCTTACCTGTATGACAGGAATGGAAGAAATCCGAAGACAGTTCGGCAAC 1687  
QY 2269 ACATTGAAGTGGATCTGATTTGGAGGCTGCGCATGTATACAGATGCAATTTGAAGAATA 2328  
DB 1688 ACATTGAAGTGGATCTGATTTGGAGGCTGCGCATGTATACAGATGCAATTTGAAGAATA 1747  
QY 2329 TTTTACTCATGTTCCAAAGAGTGGTGTGTGTGATGAAGAACTCTTACTTGTGGCTTATA 2388  
DB 1748 TTTTGTCTCATGTTCCAAAGAGTGGTGTGTGTGATGAAGAACTCTTACTTGTGGCTTATA 1807  
QY 2389 AAGAACTCTCAAGAGCTGTATGAGGTGCAGTACCACTTTCATATCTAGTAGACAGACAG 2448  
DB 1808 AAGAACTCTCAAGAGCTGTATGAGGTGCAGTACCACTTTCATATCTAGTAGACAGACAG 1867  
QY 2449 TAGTACAACTCGTGTGGACATAGTTTGGAAACAAAGTCTCTACAGAGTATCTCAGGATCTTG 2508  
DB 1868 TAGTACAACTCGTGTGGACATAGTTTGGAAACAAAGTCTCTACAGAGTATCTCAGGATCTTG 1927  
QY 2509 TAAGCATATACATCTGCCACTCTTAGGACCTTGTGCTTCTATGATGATGATTAAGCAGGC 2568  
DB 1928 TAAGCATATACATCTGCCACTCTTAGGACCTTGTGCTTCTATGATGATGATTAAGCAGGC 1987  
QY 2569 TGGTGTCTTTCAAGACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2628  
DB 1988 TGGTGTCTTTCAAGACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2047  
QY 2629 TAGTGAATATCCTTTACGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2688  
DB 2048 TGGTGGAGTACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2107  
QY 2689 GAAGAAATGGAAGTCTCTTATAGCAGGTTTATTAACCAAGATGTTAAAGTCAGACAG 2748  
DB 2108 GAAGAAATGGAAGTCTCTTATAGCAGGTTTATTAACCAAGATGTTAAAGTCAGACAG 2167  
QY 2749 AAGAAATGATGATAAGATATATCATGCTTCAGATTTGCTGATGCTTAAATGATGCCA 2808  
DB 2168 AAGAAATGATGATAAGATATATCATGCTTCAGATTTGCTGATGCTTAAATGATGCCA 2227  
QY 2809 ATAAGTTCTTGTACTTGTACTTGTACTTGTACTTGTACTTGTACTTGTACTTGTACTTGT 2868  
DB 2228 ACAAGTTCTTGTACTTGTACTTGTACTTGTACTTGTACTTGTACTTGTACTTGTACTTGT 2287  
QY 2869 TATCTACAAAGACAGGATTTGATTAACCAATATATACACTAATAGAAATGCTTC 2928  
DB 2288 TATCCAAAGACAGGATTTGATTAACCAATATATACACTAATAGAAATGCTTC 2347  
QY 2929 AGGTCTCTCATCTATATTTGGGTGAGGTTATGTTACCTGGAGTGGGAAATGTGACCAAG 2988  
DB 2348 AGGTCTCTCATCTATATTTGGGTGAGGTTATGTTACCTGGAGTGGGAAATGTGACCAAG 2407  
QY 2989 AAGAGGTACAAATGAGAGAATCATCTTGTCTTGTCTTGTCTTGTCTTGTCTTGTCTTGTCT 3048  
DB 2408 AAGAGGTATTAATGAGAGAATCATCTTGTCTTGTCTTGTCTTGTCTTGTCTTGTCTTGTCT 2467  
QY 3049 CCATTGCCAAAATTTACCTGAGAAATGAAAATATGAAATGAAATGAAATGAAATGAAATG 3108  
DB 2468 CCATTGCCAAAATTTACCTGAGAAATGAAAATATGAAATGAAATGAAATGAAATGAAATG 2527  
QY 3109 ACAAGTGGCCACATTTAAGAAACACAGGTGTATCAGGCCATGGAGTTTATGAACTAAAG 3168

DB 2528 ACAAGTGGCCACATTTAAGAAACACAGGTGTCTCGGCCATCGAGTTTATGAATGAAAG 2587  
QY 3169 ATGAATCACTGAAAGACTTCAATATGTACTTTTATCTACTTCTTCTTCTTCTTCTTCTTCTT 3228  
DB 2588 ATGAATCACTGAAAGACTTCAATATGTACTTTTACCATTTATTTCTTAAACACAGCATAGCA 2647  
QY 3229 AGGCTGAACATATGACAGAAAGAGAGAAACAAGAAACAAGAAATGAAAGATGAAAGATG 3288  
DB 2648 AGGCTGAACATATGACAGAAAGAGAGAAACAAGAAACAAGAAATGAAAGATGAAAGATG 2707  
QY 3289 CACCAACACCTTCTCAATCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTG 3348  
DB 2708 GCGCACCTTCTTCCAGAGTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTT 2767  
QY 3349 ATATCATGATGTACATCTTCTCAGGACCTTATTTGAGCGGCAATAGACAGACATTTCTAACT 3408  
DB 2768 ATGTATGATATACATCTTCTCAGGACCATCTTTGAGCGGCGAGTGGACACGGAGTCTAATC 2827  
QY 3409 TGTGACCGAAGGGATGCTTCCAAATGGCTTTTCAATATCTGCGATTTGGGTTTACTATAAG 3468  
DB 2828 TGTGACAGAGGGATGCTGCAATGGCTTCCATATATTTGCACTTGGCTTGGCTTGGCTGGAAG 2887  
QY 3469 AAGACCAACAGCTTCAAAAAGCTCTGAGAGAAAGTACATTTGACTTTTATCATAGG 3528  
DB 2888 AAGACCAACAGCTTCAAAAAGCTCTGAGAGAAAGTACATTTGACTTTTATCATAGG 2947  
QY 3529 CTTCAAGATTGGGAAGTTTCAAGCAT-----GAATATACAAATGCTTTTGGGAAAC 3579  
DB 2948 CTTCAAGATTGGGAAGTTTCAAGCATGATGCTCAGAAATATACAAATGCTTTGGGAAAGAC 3007  
QY 3580 TCAAGAGAAATTTCCAGATTAGAGGCCAGAGGACATGATTAAGTGGATTTACTTACAGATGT 3639  
DB 3008 TCAAGAGAAATTTCCAGATTAGAGGCCAGAGGACATGATTAAGTGGATTTACTTACAGATGT 3067  
QY 3640 TTGACACAGTGAAGCGATTTAGAGAAATCTTGTATTTAGTGTAGCAACCATCATAGGAT 3699  
DB 3068 TTGACACAGTGAAGCGATTTAGAGAAATCTTGTATTTAGTGTAGCAACCATCTTCAAGAC 3127  
QY 3700 CGGAATCTATTAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3759  
DB 3128 TGGAGTGGATTAAGAGTGAAGGATTTACTTATGATAAGAAAGGAGGAGGAGGAGGAGGAG 3187  
QY 3760 AAGCTGAAGCTGCTAGGCTTACATCCGAGAGATCATGCTCAGATGCTGCTGCTTACAG 3819  
DB 3188 AAGCTGAGCGCTAGGCTTACATCCGAGAGATCATGCTCAGATGCTGCTGCTTACAG 3247  
QY 3820 AAACTTCAATGAAACTCATAACTCATGATGACAAATACATGAAATGCTTGGGAAAG 3879  
DB 3248 AAACTTCAATGAAACTCATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3307  
QY 3880 AAGATTCCATTTAGGAGGAGAGACACCCAGCAGTCACTGACTACTTAGAAATGCTT 3939  
DB 3308 AAGATTCCATTTAGGAGGAGAGACACCCAGCAGTCACTGACTACTTAGAAATGCTT 3367  
QY 3940 TGGGTCTTAAAGCGGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTT 3999  
DB 3368 TGGGCTTAAAGCGGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTT 3427  
QY 4000 AAGAAAGACAGGAGGTGAAATAGAAATTAATGCTTATGCTTATGCTTCTGCTTCTGCTTCTG 4059  
DB 3428 AAGAAAGACAGGAGGTGAAATAGAAATTAATGCTTATGCTTATGCTTATGCTTCTGCTTCTG 3487  
QY 4060 AATCTACTTCTTAAACCCAGCAGGAGGAAACCCATAGAACTTCTCAGGAGAAAGCCCTAG 4119  
DB 3488 AATCCACCGCTTAAACCCAGCAGGAGGAGGCTTGGACCACTTGGGGAACCACTGG 3547  
QY 4120 ACCCACTTTTCTGATCCAGACTTGGCATATGGAATTTATACAGAAAGCTTGTGCTCATG 4179  
DB 3548 ACCCTCTTTTCTGATCCAGACTTGGCATATGGAATTTATACAGAAAGCTTGTGCTCATG 3607  
QY 4180 TAATCAGCGAGTGTGCTGGCAGAGATTTTGAAGCTGTACAGCTGTGAGCTCTCAGCAGC 4239

Db 3608 TAATGCATGCAGTGTGCTGGCAGAGTATTTTGAAGCTGTCAGCTGAGCTCGCAGCAGC 3667  
Qy 4240 GCATTCATGTTGACCTTTTGGACTTGGAAAGTGGAGAAATATCTTTGCCCTCTTTGGCAAAT 4299  
Db 3668 GCATTCACATGAGACCTGTTTGACCTGGAGAGCGCGAGTACCTATGCGCCGCTCTGCAAGT 3727  
Qy 4300 CTCCTGTGCAATAGTGTGATCCGCCCATATTCCTTTTGAACCTCAAAAGATAAAGAGTGA 4359  
Db 3728 CTCCTGTCAACACTGTCTATCCCATCCCATCTTTGACCGCGCAGAGATCAACAGTGA 3787  
Qy 4360 ATGCAGATCTCTTGTCTCAACTTTTGAACCTGTGACCGGTGGATACAGACTGTTCTGGCCA 4419  
Db 3788 ATGCGGAGCTCTTGTCTCAACTTTTGAACCTGTGACCGGTGGATACAGACTGTTCTGGCCA 3847  
Qy 4420 GAATATCAGGTATATATAAGACATGCTAAAGGAGAAACCC---CAATTCCTATTTTCT 4476  
Db 3848 GAATATCGGGTATATATAAAGCATGCTAAAGGAGAAAGCCCGCAGTTCCTGTCTGT 3907  
Qy 4477 TTAATCAAGGAATGGAGATTTCTACTTTTGGAGTTTCCATTCCTGAGTTTGGGCTTG 4536  
Db 3908 TTAATCAAGGAATGGGAGATTTCAACTTTTGAAGTTTCAATTCCTGAGTTTGGAGTTC 3967  
Qy 4537 AGTCTTCGATTAATATCAATAGCATCAAGGAAATGTTATCTCTTTGGCCACAACAA 4596  
Db 3968 AGTCTTCGTAATATATCAATAGTATCAAGGAAATGTTATCTCTTTGGCCACAACAA 4027  
Qy 4597 TTTATAGAAATGGATTTGAAGTGCCACCTGATGAAAGGGATCTCGAGTCCCATGCTGA 4656  
Db 4028 TTTACAGAAATGGCCTGAAAGTGCCCTCTGATGAACCTAGACCCACGAGTGCCTATGA 4087  
Qy 4657 COTGGAGCACCTGCGGCTTCTACTATCCAGGCAATTTGAAATCATTTGGGAGATGAAGGAA 4716  
Db 4088 COTGGAGCACCTGCGGCTTCTACCATCCAGGCAATTCGAAACCTGTTGGGAGATGAAGGAA 4147  
Qy 4717 AACCTCTGTTTGAGCACCTTCAAAATAGGCAGCATATGCTCTGAAGCATTAATCCAGT 4776  
Db 4148 AACCTCTATTTGGAGCACCTTCAAAATAGAGCATAGCGCTCTGAAGGCGCTAATGCAGT 4207  
Qy 4777 TTGCAGTTGCACAGAGGATTAACCTGTCCTCAGGCTCTGTATACAGAAACATCTGGTCTGC 4836  
Db 4208 TTGCAGTTGCACAGAGGCTTACCTGCGCTCAGGCTCTGTATACAAACATCTGGCTCGC 4267  
Qy 4837 TTCTATCAGTTGTTCTTCTTAACATATAAATAGAAGATACACCATGCCCTTCTGCTATAG 4896  
Db 4268 TCTCTGCAGTTATCTTCTTAACCTGCAATCAGAAATACACAGGCTTCTGCTGCTGG 4327  
Qy 4897 ATCTGTTTCATGTTTGGTGGGTGCTGTGTAGCATTCCTATCCATTCCTGATGGGATGACC 4956  
Db 4328 ATCTCTTCATGTTCTGCTGGCGCAGTCTTAGCGTTCCCATCTCTGTATTTGGGATGACA 4387  
Qy 4957 CTGTTGATCTGCAGCCTTCTTCAAGTTAGTCTTCTTCCCTATACACACCTTTATCTCTCCAT 5016  
Db 4388 CCGTGGATCTGCAGCGCTGCCACTTAGTCTTCTATATACACACCTTATCTCTCCATC 4447  
Qy 5017 TGATCACCATGGCACATGCTTCAGATCTACTTACAGTAGACA-----CAG 5064  
Db 4448 TGATCACCATGGGCACATGCTTCAGATCTACTTACAGTAGACA-----CAG 4507  
Qy 5065 GCCTACCCCTTGTCTCAGTTCAAGAGACAGTGAAGAGGCTCAATCCGCACTTCTTCTTCT 5124  
Db 4508 GGCGCGCTGCTGAGGGTGAAGAGATAGTGAAGAGGCTCCTGTGCATCTGCTTCT 4567  
Qy 5125 TTGCAGAAATTTCTCAATATACAAAGTGGCTCCATTTGGGTGTATATTCCTGGCTGGTAT 5184  
Db 4568 TTGTGGAAGTGTGCGACACAGACGCGCTCACTGGGTGCGGTGCTCCGCGCTGGTACC 4627  
Qy 5185 TGTGGGTCTCAGTGAAGATGGCATCACCCCTTATCTTCCCTGTGCTGATTTTCTTCC 5244  
Db 4628 TGTGGCTCTCCCTGAGGAGCGCATCACCCCTTACCTCCGCTGTGCTGCACTCTTCTTCC 4687  
Qy 5245 ACTATTTACTTGGGTAACCTCCGCTGAGGAAGTGCATACCAATTTCTGCAAGAGGAGAT 5304  
Db 4688 ACTATTTACTTGGAGTAGCTCCGCTGGAAGAACTGTTTGGCAATTTCTGCTGAAGAGAGAT 4747

Qy 5305 ACAGTGCACCTCTGAGCTATCTATCTTTACCTACAAAATTTGTTCTGCTCTTCCAGGAAT 5364  
Db 4748 TCAGTGCACCTCTGAGCTATCTATCTTTACCCACAAAATTTGTTCTGCTCTTCCAGGAAT 4807  
Qy 5365 ATTGGGATCTGTAAGGCCCTTGTCTCCAGAGTGTGTGAGATCCTGCTTACTTAAACT 5424  
Db 4808 ATTGGGATCTGTAAGGCCCTTACTACAGAGTGTGTGAGATCCTGCTTACTCAAGT 4867  
Qy 5425 GTTTGAAGCAAAAACACCGTGTGCTAGCTACCTTAGAAAAGAAAATAGTTGATAGAGC 5484  
Db 4868 CTTTGAAGCAAAAAGTGTCTGTGCTAGCTACCTTAGAAAAGAAAATAGTTGATAGAGC 4927  
Qy 5485 TTCCTGATGACTATGAGTGCCTCTCTGAATCAAGCTTCTCAATTCAGGTGCCACCGCTG 5544  
Db 4928 TTCCTGAGGACTACAGCTGTCTCTAAATCAGGCTCTCACTTTAGGTGTCCACGCTG 4987  
Qy 5545 CAGATGATGAGCAAGCAATCCTGCTCTCTGCTTTTCTGTTGGGCTATACTATGTTCTC 5604  
Db 4988 CAGATGATGAGCAAGCAATCCTGCTCTCTGCTTTTCTGTTGGGCTATACTATGTTCTC 5047  
Qy 5605 AGAATATTGCTGCCAGGAAATTTGAACGGGAGAGTGTGAGCTTGCATTTTTCAGC 5664  
Db 5048 AGAATATTGTTGCCAGGAAATTTGAATGGGAGAGTGTGAGCTGCGTTCATG 5107  
Qy 5665 CACTTCACTGTGGAGCGGAGTCTGCATTTTCTTAAATAATCAGAAATGCCAGTGGTCC 5724  
Db 5108 CGCTTCACTGTGGTGTGAGCTGTGCATTTTCTTAAATAATCCGAGATGCAGGTGGTCC 5167  
Qy 5725 TGGTTGAAGTAAAGCCAGAGGCTGTGCTATCCAGCTTCTTACTTGGATGAATAGAG 5784  
Db 5168 TGGTTGAAGTAAAGCCAGAGGCTGTGCTATCCAGCCCTTACTTGGATGAATAGAG 5227  
Qy 5785 AAACAGACCTTGGCTGGAAGGGGCAACCCCTTCTTCTGCTGAGCGGTATCGGA 5844  
Db 5228 AAACAGACCCAGGCTTAAAGAGAGAAACCCATCTTATCTCGGAGCGGTATCGGA 5287  
Qy 5845 AGCTCCATTTGGTCTGGCAACACACTGCATTTATAGAAGAGATTTCTAGGAGCCAGAGA 5904  
Db 5288 AGCTGCATTTGGTCTGGCAACAGCACTGCATTTATAGAAGAGATTTCTCGGAGCCAGAGA 5347  
Qy 5905 CTATCAGATGTTATTTGGATTCAGTGGAGTCTACTGTAGCTCCACTTCCGCTCAAG 5964  
Db 5348 CTATCAGATGCTATTTGGATTTAACTGGCAGTCTACTGTAGCTTCTGCTCTCAAG 5407  
Qy 5965 ACATCACAATGACGACAGTAGTAAAGGCTGATTCAAAATATGGAAGAACTTCTGAGG 6024  
Db 5408 ACATCAGTGTGACATCAATTAAGAGACTGATCTAAAATTTCTAGAGAACTTCTGAGG 5467  
Qy 6025 GCTGGGAAAGTATTTGGAGGCTTTTGTCTCCATGTCAGGTTCACTTACATCAATAAAT 6084  
Db 5468 ACGGGGAAGTATTTGGAGGCTTTTGTATCCATGTCAGAGTTCACACATTAATAAAT 5527  
Qy 6085 ATTTCTTAATGGAGTATTTGCTTCAATTAGCAACATATGCTTCCAGGAAAAA-AGGAC 6143  
Db 5528 ATTTCTTAATGGATATTTGCTTCAATTTATCAACATAAGCTTCAAGGAAAAAACAAG 5587  
Qy 6144 ATAGATCAATCTGT-----TTT 6160  
Db 5588 ATAGATTAATGTTTTATGTTCTAGAACACTAAAGAAATGCTTGTTCATCCAAAGTCTAT 5647  
Qy 6161 ATGCTGATGATTTCCAGGAATTTATTTCCCTTCAATAATTTGCTCATTTTATTTT 6220  
Db 5648 TTTCTGTAATTTTCCAGAAAACTTCTTCCCTTCAATAAAGTCTTCTAGTTTCAATTT 5707  
Qy 6221 CATCCACTGGTAGATGAAGTCAAGTCAACAGTGTAGACATTTTATGTTGTTGTTAAC 6280  
Db 5708 CACCCACTGGTAAATGAGGTCACATTAAGCATTTTGTGGACATTTCTCCATCTGCTAAC 5767  
Qy 6281 TCTTCTGCAATTTTGTATTTGGTGT 6307  
Db 5768 ATCTCTGCACCTTTGTATTTGGTGT 5794

```

RESULT 3
AAC86933 standard; cDNA; 6395 BP.
XX AC AAC86933;
XX DT 02-APR-2001 (first entry)
XX DE Nucleotide sequence of Mouse Ub1 protein.
XX KW Ub1; E3-type protein; ubiquitin system; ubiquitin-protein ligase;
XX KW N-end rule pathway; intracellular pathogen; Lysteria monocytogenes;
XX KW Yersinia enterocolitica; muscle wasting; infection; ss.
XX OS Mus sp.
XX FH Key Location/Qualifiers
XX FT CDS 115..5388
XX FT /*tag= a
XX FT /product= "Ub1"
XX PN US6159732-A.
XX PD 12-DEC-2000.
XX PF 11-JAN-1999; 99US-0228317.
XX PR 02-DEC-1997; 97US-0982956.
XX PA (CALY ) CALIFORNIA INST OF TECHNOLOGY.
XX PI Kwon YT, Varshavsky A;
XX FI WPI: 2001-090278/10.
XX DR P-PSDB; AAB31162.
XX CC The present sequence encodes a murine Ub1 enzyme. Ub1 is an E3-type
XX CC protein of the ubiquitin system. Specifically, it is a ubiquitin-protein
XX CC ligase. The enzyme is specific for destabilising residues exposed at
XX CC the N-terminus of protein substrates. Inhibition of the expression of
XX CC Ub1 gene in a cell results in inhibition of the N-end rule pathway.
XX CC The method is used for treatment of mammalian cells infected with an
XX CC intracellular pathogen, e.g. Lysteria monocytogenes or Yersinia
XX CC enterocolitica. Inhibition of N-end rule pathway is also useful for
XX CC treating various diseases associated with wasting of muscle tissue and
XX CC infections.
XX PS Claim 3; Column 15-28; 18pp; English.
XX CC
XX CC The present sequence encodes a murine Ub1 enzyme. Ub1 is an E3-type
XX CC protein of the ubiquitin system. Specifically, it is a ubiquitin-protein
XX CC ligase. The enzyme is specific for destabilising residues exposed at
XX CC the N-terminus of protein substrates. Inhibition of the expression of
XX CC Ub1 gene in a cell results in inhibition of the N-end rule pathway.
XX CC The method is used for treatment of mammalian cells infected with an
XX CC intracellular pathogen, e.g. Lysteria monocytogenes or Yersinia
XX CC enterocolitica. Inhibition of N-end rule pathway is also useful for
XX CC treating various diseases associated with wasting of muscle tissue and
XX CC infections.
XX SQ Sequence 6395 BP; 1802 A; 1376 C; 1519 G; 1698 T; 0 other;

Query Match 70.3%; Score 4435.4; DB 22; Length 6395;
Best Local Similarity 86.3%; Pred. No. 0;
Matches 4997; Conservative 0; Mismatches 721; Indels 69; Gaps 6;

QY 590 TTCAGGGCCGTCGTAAAGTGTGTCCTCTC-TCAGAGCGTCAAGTTCCTCAAGATGCGGACGAGG 708
DB 8 TTCAGGGCCGTCGTAAAGTGTGTCCTCTC-TCAGAGCGTCAAGTTCCTCAAGATGCGGACGAGG 67
QY 649 TGCCTCTGCGGGGGTCCGCAACTGCAGGCGTCAAGTTCCTCAAGATGCGGACGAGG 708
DB 68 AGCTGGGGGGGGGGTCCGGAACCTGCGGGGCTTCGTTCCCTTAAGATGCGGACGAGG 127
QY 709 AGGCTGGAGGTACTGAGAGGTGGAATACAGCGGAGGTATACCCAGACCCCTCAGCGTC 768
DB 128 AGATGGAGCGGCGCGAGAGGTGACGTCAGCCCGGAGCCCTCCCTTGGCCCGGACGCGC 187
QY 769 TGGCATCTTGTGGGATCAGCAAGTTGATTTTATCTGCTTTCTGTGCATCATTTGGGCAC 828

```



Db	1268	AAAACTCTTTGCTATGGAATTTGTGCAAGTATTATATAAACAACCTGCAGAAAGAGTACATCA	1327
QY	1909	GTGATGATCATGACAGAAAGTATCTCTATAAATGCACTTTTCAGTTCAGATGTTTACTGTTTC	1968
Db	1328	GGCAGCACCAGAGAGAACATCTCCATTAACCGCCCTGTCGCTGCAGATGCTCACCGTCC	1387
QY	1969	CTACTCTGGCTCGACATCTTATTAAGACAGCAAGTGTATCTCTGTCATTTACTGAAACTC	2028
Db	1388	CGACCTTGGCCCGGATCTTATTAAGAGCAGCAAGTGTATTTCTGTCATTTACTGAAACGC	1447
QY	2029	TGCTAGAAGTTTTACTCTAGTACTTGGACAGCAACAATAAATTCAACTTCCAGGGTTATA	2088
Db	1448	TGCTAGAAGTTTTACTCTAGTACTTGGACAGCAACAATAAATTCAACTTCCAGGGTTATA	1507
QY	2089	GCCAGGACAAATTTGGGAAGAGTATATGCAATATATGACCTTAAAGTATATCTCTGATCA	2148
Db	1508	GCCAGGACAAACTGGGAAGAGTCTAGCGAGTTATATGTGACCTTAAAGTATATCTCTGATTA	1567
QY	2149	GCAAAACCCACAATATGGACAGAAAGATTAAGAAATGCAAGTTTCCAGGGTTTTCATCTT	2208
Db	1568	GCAAGCCTGTCTATATGGACAGAAAGATTAAGAGCGAGTTTCCCTGGAAGGTTTCCGGTCTT	1627
QY	2209	TTTTGAAGATTTCTTACTCTGATGCAAGGAAATGGAAGAAATCCGAACAGGTTGGCAAC	2268
Db	1628	TTCTGAAGATTTCTTACTCTGATGCAAGGAAATGGAAGAAATCCGAACAGGTTGGCAAC	1687
QY	2269	ACATTCAGTGGATCTCTGATTTGGGAGCTGCCATTCCTATACAGATGCAATTCGAAGAATA	2328
Db	1688	ACATTCAGTGGACCTGACTGGAGGCTGCCATTCCTATACAGATGCAATTCGAAGAATA	1747
QY	2329	TTTTTACTCATGTTCCAAAGAGTGGTGTGTTGATGAAGAATCTTACTTTGTTGGCTTATA	2388
Db	1748	TTTTTGTCTATGTTCCAAAGAGTGGTGTGTTGATGAAGAATCTTACTTTGTTGGCTTATA	1807
QY	2389	AGAATGTCACAAAGCTGTGATGAGGTGCAGTACCAAGTTTCATATCTAGTAGCAACACAG	2448
Db	1808	AGAATGTCACAAAGCTGTGATGAGGTGCAGTACCAAAATTCATGTCCAGTACCAACAGAG	1867
QY	2449	TAGTCAATTCGTGTGGACATAGTTTGGAAACAAAGTCCCTACAGATGATCTGAGGATCTTG	2508
Db	1868	TAGTCAATTCGTGTGGATAGTCTGGAACCAAAATCCCTACAAAGTCTGAGGACCTTG	1927
QY	2509	TAAGCATACATGCTGCCACTCTTAGAACCCCTTCTGCTGCTTCTCATGTACGTTTAAAGCAGC	2568
Db	1928	TAAGCATACATGCTGCCACTCTTAGAACACTTCTGCTGCTTCTCATGTACGTTTAAAGCAGC	1987
QY	2569	TGGGTGCTGTTTCAAGACTGCATGAATTTGTGTCTTTTTCAGAGACTTTCAAGTAGAGGTAC	2628
Db	1988	TAGGTGCTATTTTCAAGACTGCATGAATTTGTGTCTTTTTCAGAGACTTTTCAAGTAGAGGTCC	2047
QY	2629	TAGTGAATATCTTTTACGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	2688
Db	2048	TGGTGGATGACCGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	2107
QY	2689	GAAGAAATGGACTGCTCTTATTAGCCAGGTGTTTATTACCAAGATGTTAAAGTGCAGAG	2748
Db	2108	GAAGAAATGGACTGCTCTCTATCAGCCAGGTGTTTATTACCAAGATGTTAAAGTGCAGAG	2167
QY	2749	AAGAAATGATGATAAAGATATCATCATGCTTTCAGATGCTGCTGCTGCTGCTGCTGCTGCTG	2808
Db	2168	AGGAAATGATGATAAAGATATCATCATGCTTTCAGATGCTGCTGCTGCTGCTGCTGCTGCTG	2227
QY	2809	ATAAGTTCTTGTACTGTTTACGATGATGATGATGATGATGATGATGATGATGATGATGATG	2868
Db	2228	ACAAGTTCTTGTACTGTTTACGATGATGATGATGATGATGATGATGATGATGATGATGATG	2287
QY	2869	TATCTACAAAACACAGGATTTGATTAACAATATATAACATTAATAGAGAAATGCTTTC	2928
Db	2288	TATCTACAAAACACAGGATTTGATTAACAATATATAACATTAATAGAGAAATGCTTTC	2347
QY	2929	AGGTCTCTATCTATTTGTTGGTGGAGCTTATGATGATGATGATGATGATGATGATGATGATG	2988
Db	2348	AGGTCTCTATCTATTTGTTGGGAGACGTTATGATGATGATGATGATGATGATGATGATGATG	2407
QY	2989	AAGAGTCAACAATGAGAGAAATCATCTACTTGTCTTTTGCATTTGAACCCATGCCACACAGTG	3048
Db	2408	AGGAGTTTATAATGAGAGAGATTTACTCTACTTTTGTCTTTTGCATTTGAACCCATGCCACACAGTG	2467
QY	3049	CCATTGCCAAAAATTTACCTCAGAAATGAANAATTAATGAACCTGGCTTAGAATGTCATAA	3108
Db	2468	CCATTGCCAAAAATTTACCTCAGAAATGAANAATTAATGAACCTGGCTTAGAATGTCATAA	2527
QY	3109	ACAAAGTGGCCACATTTAAGAAACACAGGTGATCAGGCGCATGAGCTTTATGAACCTAAAAG	3168
Db	2528	ACAAAGTGGCCACATTTAAGAAACACAGGTGATCAGGCGCATGAGCTTTATGAATGAAAG	2587
QY	3169	ATGAATCAGCTGAAAGACTTCAATATGTACTTTTATCTACTTCCAAAAACCCAGCATAGCA	3228
Db	2588	ATGAATCAGCTGAAAGACTTCAATATGTACTTTTACCATTTCTTAAACACACAGCATAGCA	2647
QY	3229	AGGCTGAACATATGCAGAAAGAGGAGAAACAAAGAAACAAAGAAATTAAGATGAAGCATTTGCCGC	3288
Db	2648	AGGCTGAACATATGCAGAAAGAGGAGAAACAAAGAAACAAAGAAATTAAGATGAAGCATTTGCCGC	2707
QY	3289	CACCACCACTCTCTGAAATTTCTGCCCTGCTTTTCAGCAAAAGTGAATTAACCTTCTCAACTGTG	3348
Db	2708	CGCACCTCTCTCAGAGTCTTCTGCCCTGCTTTTCAGCAAAAGTGAATTAACCTTCTCAACTGTG	2767
QY	3349	ATATCATGATGTACATTTCTCAGGACCGTATTTTGAGCGGCAATAGACACAGATTTCTAACT	3408
Db	2768	ATGTTATGATATACATCTCTCAGGACCATCTTTTGAGCGGCAATAGACACAGATTTCTAACT	2827
QY	3409	TGTGGCCGAGGAGTCTCCAAATGCTTTTCTATATTTCTGCTGCTGCTGCTGCTGCTGCTGCTG	3468
Db	2828	TGTGGACAGAGGAGTCTGCGATGCGTTCCTATATATTTGCGCATGCTGCGCTGCTGCTGCTG	2887
QY	3469	AGAGCAACAGCTTCAAAAAGCTCTGGAAGAGAGTAAACATTTGACATTTTATCATAGG	3528
Db	2888	AGAGCAACAGCTTCAAAAAGCTCTGGAAGAGAGTAAACATTTGACATTTTATCATAGG	2947
QY	3529	CTTCAAGATTTGGGAAGTTTACGCCAT-----GAATATACAAATGCTTTTGGAAAAAC	3579
Db	2948	CTTCAAGATTTGGGAAGTTTACGCCATGAATGCTCAGAAATATACAAATGCTCTTGGAAAGAC	3007
QY	3580	TCAAAGAAATTTCCCACTTTAGAACGCCAGAGACATGATAACGTTGGATCTTCAGATGT	3639
Db	3008	TCAAAGAAATTTCCCACTTTAGAACGCCAGAGACATGATAACATGGAATCTTCCAGATGT	3067
QY	3640	TTGACACAGTCAAGCGATTAAGAGAAATCTTCTTTTAAATTTGACCAACCATCAGAT	3699
Db	3068	TTGACACAGTCAAGCGATTAAGAGAAATCTTCTTTTAAATTTGACCAACCATCAGAT	3127
QY	3700	CGGAATCTATTAAAGATGATGAGATTTACTCATGATAAAGAAAAAGCAGAAAGAAAGAA	3759
Db	3128	TGAGTGATTAAGATGAGGAGATTTACTCATGATAAAGAAAAAGCAGAAAGAAAGAA	3187
QY	3760	AACTGTAAGCTGCTAGCTACATGCCAGAAAGATCATGGCTCAGATGCTGCTTTCACAGA	3819
Db	3188	AACTGTAAGCTGCTAGCTACATGCCAGAAAGATCATGGCTCAGATGCTGCTTTCACAGA	3247
QY	3820	AAAATCTCATTAAGCTCAATAAATCATGATGATGATGATGATGATGATGATGATGATGATG	3879
Db	3248	AAAATCTCATTAAGCTCAATAAATCATGATGATGATGATGATGATGATGATGATGATGATG	3307
QY	3880	AAATCTCATTAAGCTCAATAAATCATGATGATGATGATGATGATGATGATGATGATGATG	3939
Db	3308	AAATCTCATTAAGCTCAATAAATCATGATGATGATGATGATGATGATGATGATGATGATG	3367
QY	3940	TGGTCTCTAAACGGGCTTCTGTTACTGAAAAAGGAGGTGCTGACGCTGCATCTCTTCC	3999
Db	3368	TGGTCTCTAAACGGGCTTCTGTTACTGAAAAAGGAGGTGCTGACGCTGCATCTCTTCC	3427
QY	4000	AAGAAGAACAGGAGTGAATAATAGAAATTAATGCCATGATTAATTCGGCTGTGTGCCAGA	4059
Db	3428	AAGAAGAACAGGAGTGAATAATAGAAATTAATGCCATGATTAATTCGGCTGTGTGCCAGA	3487



Db 5648 TTCTGCTAATATTTCCAGAAACTCTTTCCCTTCATAACTGCTCTAGTTCAITTCATAT 5707  
QY 6221 CATCCACTGGTAGAGTGAAGTCAACAGCTGTAGACATTTTATGTGTTGGTTAAC 6280  
Db 5708 CACCCACCTGGTGAATGAGTCAATTAAGCATTTGTGGACATTTCTCCATCTGGCTAAC 5767  
QY 6281 TCTTCTGCAATTTTGTATTTGGTGT 6307  
Db 5768 ATCTCTGCACCTTGTATTGGTGT 5794

## RESULT 4

AAH14878  
ID AAH14878 standard; cDNA; 2550 BP.

XX AAH14878;

XX 26-JUN-2001 (first entry)

DE Human cDNA sequence SEQ ID NO:12731.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

XX Homo sapiens.

XX EP1074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-0116126.

XX 29-JUL-1999; 99JP-0248036.

XX 27-AUG-1999; 99JP-0300253.

XX 11-JAN-2000; 2000JP-0118776.

XX 02-MAY-2000; 2000JP-0183767.

XX 09-JUN-2000; 2000JP-0241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isoqai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602  
PT full-length cDNAs defined in the specification, and for the detection  
PT and/or diagnosis of the abnormality of the proteins encoded by the  
PT full-length cDNAs -

XX Claim 8; SEQ ID 12731; 2537pp + CD ROM; English.

XX The present invention describes primer sets for synthesising 5602  
CC full-length cDNAs defined in the specification. Where a primer set  
CC comprises: (a) an oligo-dr primer and an oligonucleotide complementary  
CC to the complementary strand of a polynucleotide which comprises one of  
CC the 5602 nucleotide sequences defined in the specification, where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in  
CC the specification. The primer sets can be used in antisense therapy and  
CC in gene therapy. The primers are useful for synthesising polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to  
CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632  
CC represent oligonucleotides, all of which are used in the exemplification  
CC of the present invention.

XX Sequence 2550 BP; 830 A; 454 C; 558 G; 708 T; 0 other;  
SQ Query Match 40.4%; Score 2546.8; DB 22; Length 2550;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 2548; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1189 ATCATGACCTGGAGAGCAGTACTATAAAGAGAATTCACGCTGTCGGTTGAATGAAG 1248  
Db 1 ATCATGACCTGGAGAGCAGTACTATAAAGAGAATTCACGCTGTCGGTTGAATGAAG 60  
QY 1249 AGGTAATGTCCAAAGCCAGGAAATATTTCCCTTCAGTGATAAAATATGTCGTAAGAATGA 1308  
Db 61 AGGTAATGTCCAAAGCCAGGAAATATTTCCCTTCAGTGATAAAATATGTCGTAAGAATGA 120  
QY 1309 CTATATGGGAAGAGGAAAAAGAACTGCCTCCTGAACTCCAGATAAGGGAGAAAAATGAAA 1368  
Db 121 CTATATGGGAAGAGGAAAAAGAACTGCCTCCTGAACTCCAGATAAGGGAGAAAAATGAAA 180  
QY 1369 GATACATATTGTGCTCTTTTCAATGATGAACACCATTCATATGACCACGTCATATACAGCC 1428  
Db 181 GATACATATTGTGCTCTTTTCAATGATGAACACCATTCATATGACCACGTCATATACAGCC 240  
QY 1429 TACAAAGAGCTCTTGACTGTGAGCTCGCAGAGGCCAGTTGCATACCTGCCATTGACA 1488  
Db 241 TACAAAGAGCTCTTGACTGTGAGCTCGCAGAGGCCAGTTGCATACCTGCCATTGACA 300  
QY 1489 AAGAGGTCGTCGGGCTGTTAAAGCGGAGGCTTATGCTGCTGCCAGGAAGCAAGGAAG 1548  
Db 301 AAGAGGTCGTCGGGCTGTTAAAGCGGAGGCTTATGCTGCTGCCAGGAAGCAAGGAAG 360  
QY 1549 ATATAAGAGTCATTCAGAAAAATGCTCTCAACATCCACTTCATGTAGAAATGATACACT 1608  
Db 361 ATATAAGAGTCATTCAGAAAAATGCTCTCAACATCCACTTCATGTAGAAATGATACACT 420  
QY 1609 CAGAGATTATGGCTCATCAGAAATTTGCTTCCGCTCTTGGTTCCTGGATGAACAAATTA 1668  
Db 421 CAGAGATTATGGCTCATCAGAAATTTGCTTCCGCTCTTGGTTCCTGGATGAACAAATTA 480  
QY 1669 TGAGCTATTCAAGTGACTTTAGGCAGATCTTTTGGCAAGCATGCTCTAGAGAAGAACCTTG 1728  
Db 481 TGAGCTATTCAAGTGACTTTAGGCAGATCTTTTGGCAAGCATGCTCTAGAGAAGAACCTTG 540  
QY 1729 ACTCGAGAATCCCTGCTCATAGCAGGTTAATGCTTTGGGATGCAAAAGCTTTATAAAG 1788  
Db 541 ACTCGAGAATCCCTGCTCATAGCAGGTTAATGCTTTGGGATGCAAAAGCTTTATAAAG 600  
QY 1789 GTGCCGTAAGATCCTTCATGAATGATCTTCAGCAGGTTTTTTTATGGAGATGGAATACA 1848  
Db 601 GTGCCGTAAGATCCTTCATGAATGATCTTCAGCAGGTTTTTTTATGGAGATGGAATACA 660  
QY 1849 AAAAATCTTTGCTATGGAATTTGGAAGTATTATAAACAATGGAAGAAAGATATATCA 1908  
Db 661 AAAAATCTTTGCTATGGAATTTGGAAGTATTATAAACAATGGAAGAAAGATATATCA 720  
QY 1909 GTGATGATCATGACAGAAATGATCTCTATAACTGCATTTTCAGTTCAGATGTTTACTGTTTC 1968  
Db 721 GTGATGATCATGACAGAAATGATCTCTATAACTGCATTTTCAGTTCAGATGTTTACTGTTTC 780  
QY 1969 CTACTCTGGCTCGACATCTTTATGGAAGCAGAAATGTTATCTCTGTCATTTACTGAAATCTC 2028  
Db 781 CTACTCTGGCTCGACATCTTTATGGAAGCAGAAATGTTATCTCTGTCATTTACTGAAATCTC 840  
QY 2029 TGCTAGAAATTTTACTCGAGTACTTGGACAGGAACAATAAATTCACATTCAGGCTTATA 2088  
Db 841 TGCTAGAAATTTTACTCGAGTACTTGGACAGGAACAATAAATTCACATTCAGGCTTATA 900  
QY 2089 GCCAGGACAAATTTGGGAAGAGTATATGACGTAATATGTGACCTAAAGTATATCCCTGATCA 2148  
Db 901 GCCAGGACAAATTTGGGAAGAGTATATGACGTAATATGTGACCTAAAGTATATCCCTGATCA 960  
QY 2149 GCAAAACCCACAAATATGGACAGAAAGATTAAAGAATCAGTTCCTTGAAGGTTTTTCGATCTT 2208  
Db 2149 GCAAAACCCACAAATATGGACAGAAAGATTAAAGAATCAGTTCCTTGAAGGTTTTTCGATCTT

Db 961 GCAACCCACAATATGACAGAAAGATTAGAATGCAGTTCCTTGAAGGTTTTCATCTT 1020  
Qy 2209 TTTTGAAGATTCTTACCTGTATGCAGGAATGGAAGAAATCCGAAGACAGGTGGGCAAC 2268  
Db 1021 TTTTGAAGATTCTTACCTGTATGCAGGAATGGAAGAAATCCGAAGACAGGTGGGCAAC 1080  
Qy 2269 ACATTGAAGTGGATCCTGATTGGAGGCTGCCATTGCTATACAGATGCAATTTGAAGATA 2328  
Db 1081 ACATTGAAGTGGATCCTGATTGGAGGCTGCCATTGCTATACAGATGCAATTTGAAGATA 1140  
Qy 2329 TTTTACTCATGTTTCCAAAGTGGTGTCTTGTGATGAAGAACTCTTACTTCTGCTTATA 2388  
Db 1141 TTTTACTCATGTTTCCAAAGTGGTGTCTTGTGATGAAGAACTCTTACTTCTGCTTATA 1200  
Qy 2389 AAGAAATGTCACAAAGCTGTGATGAGTGCAGTACCAGTTTCATATCTAGTAGCAAGACAG 2448  
Db 1201 AAGAAATGTCACAAAGCTGTGATGAGTGCAGTACCAGTTTCATATCTAGTAGCAAGACAG 1260  
Qy 2449 TAGTACAATCGTGTGGACATAGTTTGGAAACAAAGTCCACAGAGTATCTGAGGATCTTG 2508  
Db 1261 TAGTACAATCGTGTGGACATAGTTTGGAAACAAAGTCCACAGAGTATCTGAGGATCTTG 1320  
Qy 2509 TAAGCATACATCTGCCACTCTCTAGGACCCCTTGTGCTCTCATGTAGTCTTAAGCAGGC 2568  
Db 1321 TAAGCATACATCTGCCACTCTCTAGGACCCCTTGTGCTCTCATGTAGTCTTAAGCAGGC 1380  
Qy 2569 TGGTGTCTGTTTCAAGACTGCATCAATTTGTGCTCTTTGAGGACTTTCAGTAGAGGTAC 2628  
Db 1381 TGGTGTCTGTTTCAAGACTGCATCAATTTGTGCTCTTTGAGGACTTTCAGTAGAGGTAC 1440  
Qy 2629 TAGTGGAAATCTCTTACGTTTGTGCTGTGTTGGTGGTCCAGGTTTGTGCTGAGATGTGC 2688  
Db 1441 TAGTGGAAATCTCTTACGTTTGTGCTGTGTTGGTGGTCCAGGTTTGTGCTGAGATGTGC 1500  
Qy 2689 GAAGAAATGGACTCTCTTATTAGCCAGGTGTTTTATTACCAAGATGTTAAGTGCAGAG 2748  
Db 1501 GAAGAAATGGACTCTCTTATTAGCCAGGTGTTTTATTACCAAGATGTTAAGTGCAGAG 1560  
Qy 2749 AAGAAATGTATGATAAAGATATCATCTGTTACAGATTGGTGCATCTTTAATGGATCCCA 2808  
Db 1561 AAGAAATGTATGATAAAGATATCATCTGTTACAGATTGGTGCATCTTTAATGGATCCCA 1620  
Qy 2809 ATAAGTCTTGTGTTACTGTTACTCAGAGGTATGAACCTTGCAGGCTTTTAAACAGACCA 2868  
Db 1621 ATAAGTCTTGTGTTACTGTTACTCAGAGGTATGAACCTTGCAGGCTTTTAAACAGACCA 1680  
Qy 2869 TATCTACAAAGACAGGATTTGATTAACAATATATACACTAATAGAAAGAAATGCTTC 2928  
Db 1681 TATCTACAAAGACAGGATTTGATTAACAATATATACACTAATAGAAAGAAATGCTTC 1740  
Qy 2929 AGTCTCATCTATATTGTGGTGAGCGTATTGTACCTGGAGTGGGAAATGTGACCAAG 2988  
Db 1741 AGTCTCATCTATATTGTGGTGAGCGTATTGTACCTGGAGTGGGAAATGTGACCAAG 1800  
Qy 2989 AAGAGGTCACAATGAGAGAAATCATCTCTGTTGCAITTTGAACCCCATGCCACACAGTG 3048  
Db 1801 AAGAGGTCACAATGAGAGAAATCATCTCTGTTGCAITTTGAACCCCATGCCACACAGTG 1860  
Qy 3049 CCATTGCCAAAATTTACCTGAGAAATGAAAATATGAAACTGGCTTAGAGAAATGTCATAA 3108  
Db 1861 CCATTGCCAAAATTTACCTGAGAAATGAAAATATGAAACTGGCTTAGAGAAATGTCATAA 1920  
Qy 3109 ACAAGTGGCCACATTTAAGAAACAGGTGTATCAGGCCATGGAGTTTATGAACATAAAG 3168  
Db 1921 ACAAGTGGCCACATTTAAGAAACAGGTGTATCAGGCCATGGAGTTTATGAACATAAAG 1980  
Qy 3169 ATGAATCAGTCAAGACTTCAATATGTACTTTTATCTACTCCAAAACCCAGCATAGCA 3228  
Db 1981 ATGAATCAGTCAAGACTTCAATATGTACTTTTATCTACTCCAAAACCCAGCATAGCA 2040  
Qy 3229 AGGCTGAACATATGACGAAGAAAAGGAGAAAACAAAGATGAAGCATTTGCCGC 3288  
Db 2041 AGGCTGAACATATGACGAAGAAAAGGAGAAAACAAAGATGAAGCATTTGCCGC 2100

RESULT 5

ABQ75898  
ID ABQ75898 standard; cdNA; 1635 bp.

XX AC ABQ75898;

XX DT 17-OCT-2002 (first entry)

XX DE Human ubiquitin relative protein 46.64 cdNA.

DE Human; ubiquitin relative protein 46.64; tumour; inflammation;  
KW immunological disease; haemopathy; human immunodeficiency virus; HIV;  
KW gene, ss.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

FT CDS 55..1329

FT /tag- a

FT /product= "ubiquitin relative protein 46.64"

XX CN1339485-A.

XX PD 13-MAR-2002.

XX PF 23-AUG-2000; 2000CN-0119708.

XX PR 23-AUG-2000; 2000CN-0119708.

XX PA (BODE-) BODE GENE DEV CO LTD SHANGHAI.

XX PI Mao Y, Xie Y;

XX DR WPI: 2002-464069/50.

XX DR P-PSDB; ABQ75898.

XX PT New polypeptide-human ubiquitin relative protein 46.64 for  
PT treating malignant tumours, inflammations, immunological diseases,  
PT haemopathy and human immunodeficiency virus infection -

XX  
PS  
XX

Claim 6; Page 24-25 (disclosure); 33pp; Chinese.

The present invention discloses a new kind of polypeptide, human ubiquitin relative protein 46.64, polynucleotides for encoding this polypeptide, and a DNA recombination process to produce the polypeptide. The present invention also discloses the method of applying the polypeptide in treating various diseases, such as malignant tumours, inflammations, immunological diseases, haemopathy and human immunodeficiency virus (HIV) infection. The current sequence represents the human ubiquitin relative protein 46.64 cDNA.

XX  
SQ . Sequence 1635 BP; 529 A; 290 C; 344 G; 472 T; 0 other;

Query Match 20.8%; Score 1311; DB 24; Length 1635;  
Best Local Similarity 99.2%; Pred. No. 0;  
Matches 1317; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY	2127	GACCTAAGTATATCTGATCAGCAACACCACATATGGACAGAAATTAAGAAATGCAG	2186
DB	1	GACCTAAGTATATCTGATCAGCAACACCACATATGGACAGAAATTAAGAAATGCAG	60
QY	2187	TTCTTGAAGGTTTTCGATCTTTTGAAGATCTTACCTGTATGCAGGAATGGAAGAA	2246
DB	61	TTCTTGAAGGTTTTCGATCTTTTGAAGATCTTACCTGTATGCAGGAATGGAAGAA	120
QY	2247	ATCCGAAGACAGGTTGGGCAACACATTTGAAGTGGATCCTGATTTGGGAGGCTGCCATTTGCT	2306
DB	121	ATCCGAAGACAGGTTGGGCAACACATTTGAAGTGGATCCTGATTTGGGAGGCTGCCATTTGCT	180
QY	2307	ATACAGATGCAATTTGAAGATATTTTACTCATGTTTCCAAAGTGGTGTGCTTGTGATGAA	2366
DB	181	ATACAGATGCAATTTGAAGATATTTTACTCATGTTTCCAAAGTGGTGTGCTTGTGATGAA	240
QY	2367	GAACTCTTACTTGTGCTTATAAGAAATGTCACAAAGCTGTGATGAGTGCGAGTACCAGT	2426
DB	241	GAACTCTTACTTGTGCTTATAAGAAATGTCACAAAGCTGTGATGAGTGCGAGTACCAGT	300
QY	2427	TTCATATCTAGTAGCAAGACAGTAGTACAAATCGTGTGGACATAGTTTGGAAACAAAGTCC	2486
DB	301	TTCATATCTAGTAGCAAGACAGTAGTACAAATCGTGTGGACATAGTTTGGAAACAAAGTCC	360
QY	2487	TACAGAGTATCTGAGGATCTTGAAGCATACATCTGCCACTCTCTAGGACCCCTTGTGCT	2546
DB	361	TACAGAGTATCTGAGGATCTTGAAGCATACATCTGCCACTCTCTAGGACCCCTTGTGCT	420
QY	2547	CTTCATGTAGTTTAAAGCAGCTGGGTGCTGTTTCAAGACTGCATGAATTTGTCTTTT	2606
DB	421	CTTCATGTAGTTTAAAGCAGCTGGGTGCTGTTTCAAGACTGCATGAATTTGTCTTTT	480
QY	2607	GAGGACTTTCAAGTAGAGGTACTAGTGGAAATATCCTTTACGTTGCTCTGTGGTTGCC	2666
DB	481	GAGGACTTTCAAGTAGAGGTACTAGTGGAAATATCCTTTACGTTGCTCTGTGGTTGCC	540
QY	2667	CAGGTTTGTGCTGAGATGTGGCGAAGAAATGGACTGCTCTTATAGCCAGGTTTAT	2726
DB	541	CAGGTTTGTGCTGAGATGTGGCGAAGAAATGGACTGCTCTTATAGCCAGGTTTAT	600
QY	2727	TACCAGATGTTTAAAGTGCAGAGAAATGTATGATAAATATCATCATCTTCAGATT	2786
DB	601	TACCAGATGTTTAAAGTGCAGAGAAATGTATGATAAATATCATCATCTTCAGATT	660
QY	2787	GGTGCATCTTTAATGGATCCCAATAGTTCTTGTGTTACTGTCTTACAGAGTATGAACCT	2846
DB	661	GGTGCATCTTTAATGGATCCCAATAGTTCTTGTGTTACTGTCTTACAGAGTATGAACCT	720
QY	2847	GCCGAGGCTTTTAAACAGACCATATCTTACAAAAGCAGGATTTGATTAAACATATAAT	2906
DB	721	GCCGAGGCTTTTAAACAGACCATATCTTACAAAAGCAGGATTTGATTAAACATATAAT	780
QY	2907	ACACTAATAGAAGAAATGCTTACAGTCCCTCATATTGTTGGGTGAGCGTTATGTACCT	2966
DB	781	ACACTAATAGAAGAAATGCTTACAGTCCCTCATATTGTTGGGTGAGCGTTATGTACCT	840

QY	2967	GGAGTGGAAATGTGACCAAGAGAGGTCACAATGAGAGAAATCATTTCACTTCTGCTTTC	3026
DB	841	GGAGTGGAAATGTGACCAAGAGAGGTCACAATGAGAGAAATCATTTCACTTCTGCTTTC	900
QY	3027	ATTGAACCATGCCCCACACAGTCCCATTTGCCAAAAATTTACCTGAGAAATGAAATGAA	3086
DB	901	ATTGAACCATGCCCCACACAGTCCCATTTGCCAAAAATTTACCTGAGAAATGAAATGAA	960
QY	3087	ACTGCTTTAGAGAATGTCTATAAACAAGTGGCCACATTTTAAAGAAACCCAGGTGTATCAGG	3146
DB	961	ACTGCTTTAGAGAATGTCTATAAACAAGTGGCCACATTTTAAAGAAACCCAGGTGTATCAGG	1020
QY	3147	CATGAGTTTATGAACATAAGATGAATCACTGAAAGACTTCAATATGTACTTTTATCAT	3206
DB	1021	CATGAGTTTATGAACATAAGATGAATCACTGAAAGACTTCAATATGTACTTTTATCAT	1080
QY	3207	TACTCCAAAACCCAGCATAGCAAGCTGAACATATGCAGAAAGAGGAGAAACAAAGAA	3266
DB	1081	TACTCCAAAACCCAGCATAGCAAGCTGAACATATGCAGAAAGAGGAGAAACAAAGAA	1140
QY	3267	AACAAAGATGAAGCATTTGCCGCCACACACCTCTCTGAATTTCTGCCCTGCTTTTCAGCAA	3326
DB	1141	AACAAAGATGAAGCATTTGCCGCCACACACCTCTCTGAATTTCTGCCCTGCTTTTCAGCAA	1200
QY	3327	GTGATTAACCTTCTCAACTGTGATATCATGATGTACATTTCTCAGGACCCGTATTTGAGCGG	3386
DB	1201	GTGATTAACCTTCTCAACTGTGATATCATGATGTACATTTCTCAGGACCCGTATTTGAGCGG	1260
QY	3387	GCAATAGACACAGATTTCAACTTTGGACCCGAAGGATGCTCCAAATGCTTTTCATATT	3446
DB	1261	GCAATAGACACAGATTTCAACTTTGGACCCGAAGGATGCTCCAAATGCTTTTCATATT	1320
QY	3447	CTGGCAT 3453	
DB	1321	CAGGCAT 1327	

RESULT 6  
AAK51709  
ID AAK51709 standard; cDNA; 6840 BP.  
XX  
AC AAK51709;  
XX  
DT 06-NOV-2001 (first entry)  
XX  
DE Human polynucleotide SEQ ID NO 254.  
XX  
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
KW nervous system disorder; arthritis; inflammation; ss.  
OS Homo sapiens.  
XX  
PN WO200157190-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 05-FEB-2001; 2001WO-US04098.  
XX  
PR 03-FEB-2000; 2000US-0496914.  
PR 27-APR-2000; 2000US-0560875.  
PR 20-JUN-2000; 2000US-0598075.  
PR 19-JUL-2000; 2000US-0620325.  
PR 01-SEP-2000; 2000US-0654936.  
PR 15-SEP-2000; 2000US-0663561.  
PR 20-OCT-2000; 2000US-0693325.  
PR 30-NOV-2000; 2000US-0728422.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Tang Yt, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;



Db 1897 TAACATTTTGCAGTCAGATGTCATGTTGTGCATCATGGGAACAATTTCTGCAATGGCGTGT 1956  
QY 3392 AGACACAGATCTTAACATTTGGACCGAAGGATGCTCCAAATGCGCTTTTCATATTCGGC 3451  
Db 1957 GGAACATAATGGATATGCGCTGGTCAGAGTCCATTCGCAAGAGGTGTACATTTAAATGG 2016  
QY 3452 ATTGGGTTTACTAGAAGAGAACACAGCTTCAAAAAGCTCCTGAAG---AAGAAGTAAC 3508  
Db 2017 CATGGCAGTACAAGAGAAAACAACATTTAGAGATGTCAAGGAGAGCATGTAGTAAC 2076  
QY 3509 ATTGGACTTTTATCAATAGGCTTCAAGATTTGGGAAGTTTCAGCCATGAAT-----AT 3559  
Db 2077 ATTTACCTTCACATCAAGAGATATCAAAACCTGTGTGAAGCGCCAAAAAATTTCTCCTAGCAT 2136  
QY 3560 ACAATAGCTTTTGGAAAACTCAAGGAATTTCCCACTTAGAAGGCCAGAGGACATCAT 3619  
Db 2137 ACTAGCTATGCTGGAACATCAAAAATGCTCCCTACCTAGAAAGTCCACAAAGACATCAT 2196  
QY 3620 AACGTGGATCTTCAGATGTTTGACACAGTGAAGCGATTAAGAGAAAAATCTTCTTTAAT 3679  
Db 2197 TCGGTGGATATGAAGACTTTTAAATGCTGTAAAGATCAGGGAGA-----GTTCAAC 2250  
QY 3680 TGTAGCAACACATCAGGATCGGAATCTATTAAAGATGATGAGATTACTCATGATAAGA 3739  
Db 2251 TACCAGTCCGCTGGCAGACAGCAAGAACCATAAATGGAAGAGAGTTCAAGGGACAAGA 2310  
QY 3740 AAAAGCAGAAAGAAAAGAAAGCTGAAGCTGCTAGGCTACATCGCCAGAGATCATGGC 3799  
Db 2311 CAAAGCTGAGAGGAAGAAAAGACAGAGATTGCGACAGTGGCAGAGAAAAGATCATGGC 2370  
QY 3800 TCAGATGCTCGCCTTACAGAAAACCTTCATTGAACTCAATAACTCATGTATGACATAC 3859  
Db 2371 TCAGATGCTGAAATCAGCGGATTTTATTGATGAAAACAAGAACTCTTTACGACAGAC 2430  
QY 3860 ATCAGAAATGCCCTGGGAAAGAAATTCATTTATGGAGGAAGAGACACCCAGCAGTCAG 3919  
Db 2431 ATTAGAATGGATGCGCTCAACCTCTGCTGTTCTTGATCAI-----AGCCCTGTGGCTC 2484  
QY 3920 TGACTACTTAGAATTTGCTTTGGTCTTAACGGGGTCCATCTGTTACTGTAAGAAAGAGGT 3979  
Db 2485 AGATATGACACTTTACAGCACTGGCGCCCGCACAACTCAGGTTCTCTGAACAAGACAAAT 2544  
QY 3980 GCTGACGTGCATCCTTTGGCAAGAACAGAGAGGTGAATAAGAAATATGCCATGGT 4039  
Db 2545 CGTTAGATGATATTGTTGCAAGAGGAGCAAGAGTTAAAGTGAAGAGCGGCAATGGT 2604  
QY 4040 ATTATCGGCTGTGTCCAGAAATCTACTGCTTTAAACCCAGACAGGGGAAAACCCATAGA 4099  
Db 2605 CTGCGGACATTTGTTTCAGAGATCAACTGTATTATCAAAAACAGAGTAATTTATTC 2664  
QY 4100 ACTCTCAGGAGAGCCCTAGACCCACTTTTCATGGATCCAGACTTGGCATATGSAACTTA 4159  
Db 2665 AGATCCAGAAAA---TATGATCCATTATTTCATGACCCCTGATCTGTTGTGGAACACA 2721  
QY 4160 TACAGAAAGCTGTGGTCATGTAATGACGCAAGTGTGCTGGCAGAGTATTTTGAAGCTGT 4219  
Db 2722 CACTAGTAGCTGTGGGCATATTATGATGCCCCATTTGTTGCAAGAGTATTTTGAATCCGT 2781  
QY 4220 ACAGCTGAGCTCTCAGCAGGCAATTCATGTTGACCTTTT-----TGACTTGA 4267  
Db 2782 TCAAGCTAAGAACAGGAGGCAACAGAGATTCAGCTATACATACGAGCTATGATGTAGA 2841  
QY 4268 AAGTGAGATATCTTTGGCCCTCTTGGCAATCTCTGTGCAATCTGTGATCTGATCCCAATTAT 4327  
Db 2842 AAACGGAGAAATCTTTGGCCCTTTGTGAACTGTTGAGTAATCTGTTATTC----- 2895  
QY 4328 TCTTTGCAACCTCAAAAGATAAACAGTGAAGATGCAAGTGCCTTGTGCTCAACTTTTGAC 4387  
Db 2896 TCTGCTGCTTCTCCCAAGAAATATTTTAAACAACAGGTTAAATTTTTCAGACCAACCA 2955  
QY 4388 CTGGCAGCGTGGATACAGACTGTTCTGGCCAGAAATATCAGGTTTAAATATAGACATGC 4447  
Db 2956 TCTGACTCAGTGGATTAGAACATAATCTCAGCAAAATAAAAGCATTTACAGTTTCTTAGGAA 3015

QY 4448 TAAAGGAGAAAACCCAAATTCCTATTCTTTTAATCAAGGAATGGGAGATCTTACTTTGGA 4507  
Db 3016 AGAAGAAAGTACTCTCTAATAATGCGCTCTACAAAGAAATTCAGAAAATGTGGATGAATACA 3075  
QY 4508 GTTCATTTCCATCTCTGAGTTTGGGCTTGAGTCTTCGATTAATAATTTAAATAGCATCAA 4567  
Db 3076 GCTCCTGAAGGTTTCAGGCCGTGATTTTCGCTCTAAGATCCCTTATCTTGAGAGCATAAA 3135  
QY 4568 GGAATGTTTATCTCTTTGGCCACAAACAATTTATAGAATTTGGAATGGAAGTGGCCACTGA 4627  
Db 3136 AGAATGCTAACGACATTTGGAACTGCTACCTACAAGTGGGACTAAAGGTTTCATCCCAA 3195  
QY 4628 TGAAGGATCCTCGAGTCCCATGCTGACCTGGAGCACCTCGGCTTTCACTATCCAGGC 4687  
Db 3196 TGAAGAGATCCTCGTGTTCCTCAATATGTTGGGGTAGCTGCGGTACACCATCCAAAG 3255  
QY 4688 AATTGAAAATCTATTGGGAGATGAAGAAAACCTCTGTTGGAGCACTTCAAAATAGGCA 4747  
Db 3256 CATAGAAAGAAATTTGAGTGATGAGATGAAGATTAACCATTTGTTGGTCTTTACCTTGCAGACT 3315  
QY 4748 GCATAATGGTCTGAAAGCATTAATGCAGTTTGCAGTTGCACAGAGGATTAACCTGTCTCA 4807  
Db 3316 GGATGACTGCTCTTAGGTCAATTGACGAGATTTGCGCGACGACACTGGACAGTGGCATCAGT 3375  
QY 4808 GGTCTGTATCAGAAACATCTGGTTCCTTCTATCAGTGTGTTCTTCTTAACATAAATC 4867  
Db 3376 TTCAGTGTGTGAAGGACATTTTGTAAACCTTTTGCACTCAGTGGTCCCTTAATGACAGCCA 3435  
QY 4868 AGAAGATACACCATGCGCTTCTGCTATAGATCTGTTTCATGTTTGGTGGGTGCTGTGT 4927  
Db 3436 TGAGGAATCCATGCATATTAGATATTGACATGTTTCATTTATTTGGTGGGTGTTGTCT 3495  
QY 4928 AGCATTCCTCATCTTGTATTGGGATGACCTCTGATCTGACGCTTCTTTCAGTTAGTTTC 4987  
Db 3496 TGCAATTCCTGGTTCAGTG-----TCAGGATTTTCAGGGATCAGCCT 3540  
QY 4988 TTCTATAACACACCTTTTATCTCTTCCATTTGATCACCATGCGACACATGCTTCAGATACT 5047  
Db 3541 TGGCACTGGAGCCCTTCACATTTTCCATCTGCTTACTATGCGACACATCATACAGATCTT 3600  
QY 5048 ACTTACAGTAGACACAGGCCCTAC---CCTTCTCAGGTTCAAGAGACAGACAGGAGGC 5104  
Db 3601 ACTTACCTCATGTACAGAAAGAAATGGCATGGATCAAGAAAATCCCCCTTTGGAAGAAGA 3660  
QY 5105 TCATTCCCATCTCTTTCTTTCAGAAATTTCTCAATATACAAGTGGCTCCATTTGGTG 5164  
Db 3661 ATCAGCAGTCTTCTGTTGTATAAACACTTCACCAGTATACGGGAAGTGCCTTTGAAGA 3720  
QY 5165 TGATATTCTCGCTGGGTATTTGTGGGTCTCACTGAAGAATGCGATCACCCCTTATCTTCG 5224  
Db 3721 AATACCATCCGCTGGCATCTGTGGAGGAGTGTCAAGCTGGAATCATGCTTTCCTGAA 3780  
QY 5225 CTGTGCTGCATTTGTTTCCACTATTACTTGGGGTAACTCGCGCTGAGAACTGCATAC 5284  
Db 3781 GTGTCTGCTTTATTTTTCATTTACTTAAATGGAGTTCTTCCCAACCGGACATTCAGT 3840  
QY 5285 CAATTCTCAGAAAGGAGAGTACAGTGCACCTGTAGCTATCTATCTTTACCTACAAATTT 5344  
Db 3841 ---TCCTGGAAACAGCCATTTTGAACATTTATGTAGTATCTTTCCCTACCAAAACCT 3897  
QY 5345 GTTCTGCTCTTCCAGGAATATTGGATACTGTAAAGGCCCTTTGCTCCAGAGGTGGTGTGC 5404  
Db 3898 CATTTGCCCTTTTCAAGAAAATAGTGAGTAATGAATTCACATGATTAAGTGGTGGCGC 3957  
QY 5405 AGATCCTGCCCTTAACTGTTTGAAGCAAAAAACACCGGTGGTGCAGTACCCCTAGAAA 5464  
Db 3958 TAACAGTGAAGTAAAGATATCTAGAAGGTGAAGAGATGCTATAAGATATATCAAGAGA 4017  
QY 5465 AAGAAATAGTTTATAGAGCTTCTTGATGACTATAGTGCCTTCCCTGAACTCAAGCTTCTCA 5524  
Db 4018 ATCAACAAATTAATAAACCTTCCAGAGGATTTACAGCAGCCTCATTAATCAAGCATCAA 4077





Db 937 GTCTCGGTGACTGATGCACTGTCATGGTGGTTATATCTGATGGTGAACAGCCCAATCACAC 996  
QY 2452 TACAATCGTGGACATAGTTTGGAAAAAGTCCCTACAGAGTATCTGAGGATCTGTGTAA 2511  
Db 997 TAAGCATTTTGTGACATTCAGTGGAACTATCAGATACTGTGTTTCCCAAGAAAAAGTTA 1056  
QY 2512 GCATACATCGCCACTCTCTAGGACCCCTGCTGGCTTTCATGTACGTTTAAAGCAGGCTGG 2571  
Db 1057 GCATTCACCTCCCGAGTTTCTGCTTACTTCAGGTTTACATGTATTATTAAAGCAAAAAGTG 1116  
QY 2572 GTGCTGTTTCAAGACTGCATGAATTTGTGCTTTTGGAGGACTTTCAAGTAGAGTACTAG 2631  
Db 1117 AAGTGGCATATAATTTCCAGAGCTCTACTCTTAAGTGAACCTTAGCCCAATGTTGA 1176  
QY 2632 TGAATATATCCCTTTACGTTGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 2691  
Db 1177 TAGAACACCCCTCTTAGATGCTTGTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1236  
QY 2692 GAAATGGACTGCTCTCTTATTAGCCAGGTGTTTATTACCAAGATGTTAAAGTGCAGAGAG 2751  
Db 1237 GAAATGGGTCTCTCTAGTAACCCAGATTTATTACTACCAATATGTGAATGTGCAAGAGCTG 1296  
QY 2752 AAATGTATGATAAGATATCATCTTCAGATGTGTCATCTTTAAATGATGCCCAATA 2811  
Db 1297 AGATGTTGACAAGGATGTAGTATGCTTCAGACAGGTGCTCCATGATGGATCCAAATC 1356  
QY 2812 AGTCTTGTGTTACTTCTCAGAGGTATGAATTCGCGAGGCTTTTAAACAGACCATAT 2871  
Db 1357 ATTTCTGTATGATCATCTCAGCGCTTTTGAATTTTATCAGATTTTCAGTACTCCAGACT 1416  
QY 2872 CTACAAAAG-----ACGAGGATTTGATTTAAACAATATATATACAC 2910  
Db 1417 ATGGAATAAGATTTAGTTCTGAGATTAACCAATAGGATGTTGTTTCAGCAGAACCAATACTC 1476  
QY 2911 TAATAGAAATAATCTCAGTCTCATCTATATGTTGGGTCAGCGTTATCTACTCGGAG 2970  
Db 1477 TAATAGAAATAATGCTATACCTCATATATATGCTGTTGGAGAGAGATTTAGTCTCGGAG 1536  
QY 2971 TGGGAATGTGACCAAGAGAGGTCAATATGAGAGAAATCAATTCACCTGCTTGTGCAATG 3030  
Db 1537 TTGGACAGTGAATGCTACAGATGAATCAAGCAGAGATTTATCCATCAGTTGAGTATCA 1596  
QY 3031 AACCCATGCCACAGTGCATTTGCCAAAAATTTACCTGAGATGAATAATGAAACTG 3090  
Db 1597 AGCCTATGGCTCATAGTGAATTTGGTAAAGTCTTTACCTGAGATGAGAACAAAGAGACTG 1656  
QY 3091 GCTTAGAGATGTCATAAACAAAGTGGCCACATTTAAGAACACAGGTGTATCAGGCCATG 3150  
Db 1657 GCATGGAGGTGAATCGAAGCAGTTGCCATTTTCAAGAAACCTGGATTTAAGAGGAGAG 1716  
QY 3151 GAGTTTATGAATAAAAGATGAATCACTGAAAGACTTCAATATGCTACTTTTATCATTTACT 3210  
Db 1717 GCATGATGAATGAAACCCAGATGTCGCAAGAGTTCACTGTTATTTCTATCATTCTT 1776  
QY 3211 CCAAAACCCAGCATPAGCAAGGCTGAACATATGCAGAGAAAGAGAGAGAAACAAAGAAACA 3270  
Db 1777 CAAGGGCAGAACAGTCCAGGCAGAGAGCGCAACGGAATTTGAAAGACAAATAAGAG 1836  
QY 3271 AGATGAAGCATTTGCCGCCAGCACCACTCGTGAATTTCTGCCCTGCTTTTACAGAAAGTGA 3330  
Db 1837 AAGATAGACACTCCCACTCCGCTGTTGGCTTCCATTTCTGCCCTGCTGTTTCAAGCCTGG 1896  
QY 3331 TTAACCTTCACTGATGATCATGATGATATCTTCAGGACCGTATTTTCAGGGGGCAA 3390  
Db 1897 TTAACATTTTGCAGTCAGATGTCATGTTGTGTCATCTGGGAACAATTCGCAATGGGCTG 1956  
QY 3391 TAGACACAGATTTCTAACTTGTGGACCGAAGGAGTGTCCCAATGGCTTTTTCATATTTCTGG 3450  
Db 1957 TGGACATATATGGATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2016  
QY 3451 CATTTGGGTTTACTAGAGAGAGAGCAACAGCTTCAAAAAGCTTCCTGGAAG---AAGAAGTAA 3507  
Db 2017 GCATGGCACTACAAGAGAAAAACAACATTTTAGAGATGTCAAGGAGAGCATGTAGTAA 2076

QY 3508 CATTTGACTTTTATCATTAAGGCTTCAAGATTTGGGAAGTTTCAGCCATGAAT-----A 3558  
Db 2077 CATTTACCTTCTACTCAGAAAGATATCAAAAACCTGGTGAAGCGCCAAAATAATTTCTCTAGCA 2136  
QY 3559 TACAATGCTTTTGGAAAAAATCTCAAGGAATTTCCCCAGTTTGAAGGCCAGAGAGCATGA 3618  
Db 2137 TACTAGCTATGCTGGAAACACATACAAAATGCTCCCTACCTAGAGTTCACAAAGACATGA 2196  
QY 3619 TAACGTGATACTTCCAGATGTTTGACACAGTGAAGCGATTAAAGAGAAAAATCTTGTGTTAA 3678  
Db 2197 TTCGTGGATATTGAAGACTTTTAAATGCTGTTTAAAGATGAGGGAGA-----GTTTCA 2250  
QY 3679 TTGTAGAACACATCATAGGATCGGAATCTATTAAAGAAATGATGAGATTTACTCATGATAAG 3738  
Db 2251 CTACCAGTCCCGTGCAGAGACAGAGGAACCAATAATGGAAGAGAGTTTCAAGGGACAAAG 2310  
QY 3739 AAAAGACAGAACGAAAAAGAAAGCTGAAGCTGTAGGCTTACATCCCGACAGACATCATGG 3798  
Db 2311 ACAAAGCTGAGAGGAAGAGAAAGCAGAGATTGCCAGACTGCCGACAGAAAGATCATGG 2370  
QY 3799 CTCAGATGCTGCTTACAGAAAAAATCTTCAATTTGAAACTCATAACTCATGTATGACAAATA 3858  
Db 2371 CTCAGATGCTGAAATGCGAGCGCATTTTATTGATGAACAACAAGAACTCTTTTCAGCAGA 2430  
QY 3859 CATCAGAAATGCTGGGAAAGAGATTCATTTATGGAGGAGAGACACCCCGAGCAGTCA 3918  
Db 2431 CATTAGAACTGGATGCTTCAAGCTCTGCTGCTTCTTGTATCAT-----AGCCCTGTGCTT 2484  
QY 3919 GTGACTACTCTAGAAATGCTTTGGGTCCTTAAGCGGGTCCATCTGTCTACTGAAAAGGAGG 3978  
Db 2485 CAGATATGACACTTACAGCACTGGGCCCCGCAAACTCAGGTTCTTGAAACAAGACAAT 2544  
QY 3979 TGCTGACGTGCTCTTTCGCAAGAGAACAGAGAGTGAATAATAGAAAATAATAGCCATGG 4038  
Db 2545 TCGTTTACATGTATTTGTCTCAAGAGAGCAAGAACTTAAAGTGGAAAGCAGGCAATGG 2604  
QY 4039 TATTATCGGCTGTGTCAGAAATCTACTGCTCTTAACCCAGACACAGGGGAAAAACCCATAG 4098  
Db 2605 TCTTGGCAGCATTTGTTTCAGAGATCAACTGTATTATCAAAAAACAGAAAGTAAATTTATTC 2664  
QY 4099 RACTCTCAGGAGAACCCCTAGACCCACTTTTCTCATGATCCAGACTTGGCATATAGGAATTT 4158  
Db 2665 AAGATCCCAAAAA---TATGATCCATTTATCATGCCCTGATCTGTCTGTGGAAACAC 2721  
QY 4159 ATACAGGAAGCTGTGCTCATGTAATGCACGCACTGTCTGGCAGAGAGTATTTTGAAGCTG 4218  
Db 2722 ACATGTAGTGTGGGCACATTAATGATGCCCATTTGTTGGCAAGGATTTTGTATTCG 2781  
QY 4219 TACAGCTGAGCTCTCAGCAGCGCATTTTCTGTTGACCTTTT-----TGACTTGG 4266  
Db 2782 TTCAAGCTAAAGAACAGCGGAAGCAACAGAGATTACGCTTATACATGATGATGATAG 2841  
QY 4267 AAGTGGAGAAATATCTTGGCCCTCTTTGCAAAATCTCTGTGCAATATCTGTGATCCTCCATTA 4326  
Db 2842 AAAAGGAGAAATTTCTTGGCCCTTTTGTGAATGCTTGTGATTAATCTGTTATTTTCC----- 2896  
QY 4327 TTCTTTTCAACTCAAAAGATAAACAGTGAAGATGCAGATGCTTGTGCTCAACTTTTGA 4386  
Db 2897 -TCTGCTGCTTCCAGAAATATTTTAAACACAGGTTAAATTTTTCAGACCAACCAA 2955  
QY 4387 CCCTGCGCAGGGTACAGACTGTTCTGGCCCAAGATATCAGGTTTATATATATAGACATG 4446  
Db 2956 ATCTGACTCAGTGGATTAGAACAAATATCTCAGCAATAAAAGCATTTACAGTTTCTTAGGA 3015  
QY 4447 CTAAGAGAAAAACCAATTCCTATTTTCTTAAATCAAGGAATGGGAGATTTCTACTTTGG 4506  
Db 3016 AAGAAGAAAGTACTCTCTAATAATGCTCTACAAAGAAATTCAGAAAAATGTGGATGAATTC 3075  
QY 4507 AGTTCCATTCCTCTGAGTTTGGGCTTTCGATTTTAAATATTAATTTCAATAGCATCA 4566  
Db 3076 AGCTCCCTGAAGGGTTACGGCCTGATTTTTCGCTCTAAGATCCCTTATTTCTGAGAGCATAA 3135



```
QY 2799 ATGGATCCCAATAGTTCCTTGTACTGTTACTGAGGATATGAACCTTGCAGGCTTTT 2858
Db 1 ATGGATCCCAACAAAGTTCCTTGTACTGTTACTGAGGATATGAACCTTGCAGGCTTTT 60
QY 2859 AACAGACCAATATCTACAAAGACCAAGATTTGATTAACAATATATACACTAATAGAA 2918
Db 61 AACAGACCAATATCTACAAAGACCAAGATTTGATTAACAATATATACACTAATAGAA 120
QY 2919 GAAATGCTTCAGTTCCTCATCTATATTTGTTGGTGAGGCTTATGTACCTGGAGTGGGAAT 2978
Db 121 GAAATGCTTCAGTTCCTCATCTATATTTGTTGGTGAGGCTTATGTACCTGGAGTGGGAAT 180
QY 2979 GTGACCAAGAAGAGGTCACAATGAGAGAAATCAATTCACCTTGCATTTGAACCCCATG 3038
Db 181 GTGACCAAGAAGAGGTCACAATGAGAGAAATCAATTCACCTTGCATTTGAACCCCATG 240
QY 3039 CCACACAGTCCCATTTGCCAAAATTTACCTGAGAAATTAATGAACCTGGCTTAGAG 3098
Db 241 CCACACAGTCCCATTTGCCAAAATTTACCTGAGAAATTAATGAACCTGGCTTAGAG 300
QY 3099 AATGTCTATAACAAGTGGCCACATTTAAGAAACCCAGTGTATCAGGCCATGGAGTTTAT 3158
Db 301 AATGTCTATAACAAGTGGCCACATTTAAGAAACCCAGTGTATCAGGCCATGGAGTTTAT 360
QY 3159 GAACCTAAAGATGAATCACTGAAGAGCTTCAATATGTACTTTTATCATTTACTCCAAAACC 3218
Db 361 GAACCTAAAGATGAATCACTGAAGAGCTTCAATATGTACTTTTATCATTTACTCCAAAACC 420
QY 3219 CAGCATAGCAAGGCTGAACATATGTCAGAGAAAGAGAGAAACAAAGAAAGATGAA 3278
Db 421 CAGCATAGCAAGGCTGAACATATGTCAGAGAAAGAGAGAAACAAAGAAAGATGAA 480
QY 3279 GCATTGCCGCCACACACCTCTGTAATTCCTGCTTTCAGCAAGTGAATTAACCTT 3338
Db 481 GCATTGCCGCCACACACCTCTGTAATTCCTGCTTTCAGCAAGTGAATTAACCTT 540
QY 3339 CTCAACTGTGATATCATATGATGATCATCTCAGGACCGTATTTAGCGGGCAATTAAGACACA 3398
Db 541 CTCAACTGTGATATCATATGATGATCATCTCAGGACCGTATTTAGCGGGCAATTAAGACACA 600
QY 3399 GATTCTAACTTTGGACCGCAAGGATGCTCCAAATGCTTTTCATATTTCTGGCATTTGGT 3458
Db 601 GATTCTAACTTTGGACCGCAAGGATGCTCCAAATGCTTTTCATATTTCTGGCATTTGGT 660
QY 3459 TTACTAGAGAGAGCAACAGCTTCAAAAAGCTCCTCAAGAGAAAGTAACTTTGACTTT 3518
Db 661 TTACTAGAGAGAGCAACAGCTTCAAAAAGCTCCTCAAGAGAAAGTAACTTTGACTTT 720
QY 3519 TATCATAAAGCTTCAAGATTGGGAAGTTCAGCCATGAATATACAAATGCTTTTGGAAAAA 3578
Db 721 TATCATAAAGCTTCAAGATTGGGAAGTTCAGCCATGAATATACAAATGCTTTTGGAAAAA 780
QY 3579 CTCAAAGGAATTTCCCAAGTTAGAGGCCAGAGACATGATAAGCTGATATCTCAGATG 3638
Db 781 CTCAAAGGAATTTCCCAAGTTAGAGGCCAGAGACATGATAAGCTGATATCTCAGATG 840
QY 3639 TTTGACACAGTGAAGCGATTAGAGAAAAATCTTTTAATTTAGTACCAACCATCAGGA 3698
Db 841 TTTGACACAGTGAAGCGATTAGAGAAAAATCTTTTAATTTAGTACCAACCATCAGGA 900
QY 3699 TCGGAATCTATTAGAGATGATGATGATTTACATGATAAGAAAAAGCAGACGAAAAAGA 3758
Db 901 TCGGAATCTATTAGAGATGATGATGATTTACATGATAAGAAAAAGCAGACGAAAAAGA 960
QY 3759 AAAGCTGAAGCTGCTAGGCTACATCGCCAGAAGATCATGGC 3799
Db 961 AAAGCTGAAGCTGCTAGGCTTCATCGCCAGAAGATCATGGC 1001.
```

RESULT 9  
AAC86934  
ID AAC86934 standard; cdna; 1001 BP.

```
XX AAC86934;
AC 02-APR-2001 (first entry)
DT Nucleotide sequence of a partial human Ubql protein.
XX
DE Ubql; E3-type protein; ubiquitin system; ubiquitin-protein ligase;
KW N-end rule pathway; intracellular pathogen; Lysteria monocytogenes;
KW Yersinia enterocolitica; muscle wasting; infection; ss.
OS Homo sapiens.
XX
Key Location/Qualifiers
FT 1..999
CDS /*tag= a
ET /product= "Ubql"
FT /note= "partial sequence"
XX
US6159732-A.
XX
PD 12-DEC-2000.
XX
PF 11-JAN-1999; 99US-0228317.
XX
PR 02-DEC-1997; 97US-0982956.
XX
PA (CALY ) CALIFORNIA INST OF TECHNOLOGY.
XX
PI Kwon YT, Varshavsky A;
XX
DR WPI: 2001-090278/10.
DR P-PSDB; AAB31163.
XX
PT Inhibiting the N-end rule pathway in mammalian cells for treating
PT infections and various diseases associated with muscle tissue wasting,
PT by inhibiting the expression of Ubql gene
XX
PS Claim 4; Column 27-30; 18pp; English.
XX
CC The present sequence encodes a partial Ubql enzyme. Ubql is an E3-type
CC protein of the ubiquitin system. Specifically, it is a ubiquitin-protein
CC ligase. The enzyme is specific for destabilising residues exposed at
CC the N-terminus of protein substrates. Inhibition of the expression of
CC Ubql gene in a cell results in inhibition of the N-end rule pathway.
CC The method is used for treatment of mammalian cells infected with an
CC intracellular pathogen, e.g. Lysteria monocytogenes or Yersinia
CC enterocolitica. Inhibition of N-end rule pathway is also useful for
CC treating various diseases associated with wasting of muscle tissue and
CC infections.
XX
SQ Sequence 1001 BP; 363 A; 186 C; 205 G; 247 T; 0 other;
```

```
Query Match 15.8%; Score 996.2; DB 22; Length 1001;
Best Local Similarity 99.7%; Pred. No. 1.1e-261;
Matches 998; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 2799 ATGGATCCCAATAGTTCCTTGTACTGTTACTGAGGATATGAACCTTGCAGGCTTTT 2858
Db 1 ATGGATCCCAACAAAGTTCCTTGTACTGTTACTGAGGATATGAACCTTGCAGGCTTTT 60
QY 2859 AACAGACCAATATCTACAAAGACCAAGATTTGATTAACAATATATACACTAATAGAA 2918
Db 61 AACAGACCAATATCTACAAAGACCAAGATTTGATTAACAATATATACACTAATAGAA 120
QY 2919 GAAATGCTTCAGTTCCTCATCTATATTTGTTGGTGAGGCTTATGTACCTGGAGTGGGAAT 2978
Db 121 GAAATGCTTCAGTTCCTCATCTATATTTGTTGGTGAGGCTTATGTACCTGGAGTGGGAAT 180
QY 2979 GTGACCAAGAAGAGGTCACAATGAGAGAAATCAATTCACCTTGCATTTGAACCCCATG 3038
Db 181 GTGACCAAGAAGAGGTCACAATGAGAGAAATCAATTCACCTTGCATTTGAACCCCATG 240
```



Db 488 GCATGAGAGTGTAATCGAAGCAGTTGCCCATTTTCAAGAAACCTGGATTAAACAGCAG 547  
Qy 3151 GAGTTTATGAACATAAAGATGAATCACTGAAAGACTTCATATATGTACTTTTATCATACT 3210  
Db 548 GCATGATGAACATGAACACCAAGATGTGCCAAGAGTTCAACTGTATTTCTATCATCTTT 607  
Qy 3211 CCAAAACCCAGCATAGCAGGCTGAACATATGAGAAAGAAAGAGAGAAACAAGAAACA 3270  
Db 608 CRAAGGCAGACAGTCCAGGCGAGAGACGCAACGGAATTTGAAAAGACAAATAGAG 667  
Qy 3271 AGATGAAGATTTGCCGCCACCAACCACTCTCTGAAATCTGCCCTGCTTTTCAGCAAAAGTGA 3330  
Db 668 AAGATACAGCACTCCCAACCTCCGCTGTGCTCCTCATCTCTGCTGCTGCTGCTGCTGCTG 727  
Qy 3331 TTAACCTTCTCACTGTATATCATATGATGATGATGATGATGATGATGATGATGATGATG 3390  
Db 728 TTAACATTTTGCAGTCAGATGTCATGTTGTCATGTCATGTCATGTCATGTCATGTCATG 787  
Qy 3391 TAGACACAGATCTTAACCTTGTGACCGAAGGATGCTCCAAATGGCTTTTCATATTTCTGG 3450  
Db 788 TGAACATATGATATGCTGTGTCAGAGTCCATGCTGCAAGGGTGTACATTTAATG 847  
Qy 3451 CATTTGGTTTACTAGAAAGAACCAACAGCTTCAAAAGCTCCTCGAAG---AAGAAGTAA 3507  
Db 848 GCATGCACTACAAGAGAAACAAACATTTAGAGATGTCACGGAAGACATGTAGTAA 907  
Qy 3508 CATTTGACATTTATCATAGAGCTTCAAGATTTGGAAGTTCAGGCATGAAT-----A 3558  
Db 908 CATTTACCTTCACTCAGAAAGATATCAAAACCTGCTGAAGGCCCAAAATTTCTCTAGCA 967  
Qy 3559 TACAATGCTTTTGGAAAACCTCAAGAGAAATTTCCCAAGTTAGAGGCCCAAGACATGA 3618  
Db 968 TACTAGCTATGCTGGAACCACTACAATAATGCTCCCTACCTAGAGTCCCAAGAGACATGA 1027  
Qy 3619 TAACGTGGATCTTCAGATGTTTGACACAGTGAAGCGATTAAAGAAAAATCTTGTTTAA 3678  
Db 1028 TTCGGTGGATTTAGAGACTTTTAATGCTGTTAAAAGATGAGGGAGA-----GTTTAC 1081  
Qy 3679 TTGTAGCAACCAATCAGGATCGGAATCTATTAAAGATGATGATGATGATGATGATGATG 3738  
Db 1082 CTACCAAGTCCCGTGGCAGACAGAGAGAAACCAATATGGAAGAGAGTTCAAGGGACAAAG 1141  
Qy 3739 AAAAGCAACGAAAAAGAAAGCTGAAGCTGAGCTGCTAGGCTACATCGCCAGAAATATGG 3798  
Db 1142 ACAAAAGCTGAGAGAGAGAAAGACAGAGATTTCCAGACTGCGCAGAGAAAGATCATGG 1201  
Qy 3799 CTCAGATGCTGCCTTTACAGAAAACTTCATTGAAACTATAAATCATGATGATGATGATG 3858  
Db 1202 CTCAGATGCTGAATGCGAGCGCATTTTATTGATGAAACAAAGAACTCTTTTCAGCAGA 1261  
Qy 3859 CATCAAAATGCTTGGGAAGAGAGATTCCATTATGGAGGAAGAGACACCCAGCACTCA 3918  
Db 1262 CATTAGAACTGGATGCTCAACCTCTGCTGCTTCTGATCA-----TAGCCCTGTGGCTT 1315  
Qy 3919 GTGACTACTCTAGAATTTGTTGGTTCCTTAAAGGGGTCATCTGTTACTGAAAGAGG 3978  
Db 1316 CAGATATGACACTTACAGCACTGGGGCGGCACAACTCAGGTTCTGTAACAAAGACAAT 1375  
Qy 3979 TCGTCACTGTCATCTTTGCCAAGAGAACAGAGAGGTGAAAATAGAAAATTAATGCCATGG 4038  
Db 1376 TCGTTACATGATATTGTTGTCAGAGAGCAAGAAGTTAAAGTGAAGCAGGCGCAATGG 1435  
Qy 4039 TATTATCGCCCTGTGTCCAGAAATCTPACTGCTTAAACCAAGCAGCAGGGAACCCCATAG 4098  
Db 1436 TCTTGGCAGCATTTGTTGAGAGATCAACTGTATTATCAAAAACAGAAATTTATTTATTC 1495  
Qy 4099 AACTCTCAGAGAACCCCTAGACCCACTTTTTCATGATCCAGACTTTGGCATATGGAACCT 4158  
Db 1496 AAGATCCAGAAAAATAT---GATCCATTTATCATGCAACCTGATCTGCTTTGTGGAACAC 1552  
Qy 4159 ATACAGGAAGCTGTGGTCAATGTAATGCACGCACTGTGCTGGCAGAGATATTTTGAAGCTG 4218  
Db 1553 ACACATAGCTGTGGGCACATTTATGCAATGCCATGCCCATTTGTTGGCAAGGATTTTGTTCGG 1612

Qy 4219 TACAGCTCAGCTCTCAGCAGCGCATTCATGTTTGACCTTTT-----TGACTTGG 4266  
Db 1613 TTCAAGCTTAAAGAACAGCGAAGCAACAGAGATTACGCTTATACATACAGCTATGATGTAG 1672  
Qy 4267 AAAGTGGAGAAATATCTTTTCCCTCTTTTGCAAATCTCTGTGCAATCTGTGATCCCATTA 4326  
Db 1673 AAACGGAGAAATCTCTTTGCCCTTTTGTGAATGCTTGTGATGATGATGATGATGATGATG 1732  
Qy 4327 TTCTTTTGCACCTCAAAAGATAACAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 4386  
Db 1733 TGCCTT-----CTCCAAGAAATATTTTAAACAACAGGTTAAATTTTTCAGACCAACCAA 1786  
Qy 4387 CCCTGGCAGGTTGGATACAGACTGTTCTGCCCAGATATATCAGGTTATATATATATATAT 4446  
Db 1787 ATCTGACTCAGTGGATTAGAACAAATCTCAGCAATAAAGACATTTACAGTTTCTTAGGA 1846  
Qy 4447 CTAAGGAGAAAAACCAATTCCTATTTTCTTTTAAATCAAGAAATTCAGAAATGTTGGAT 4506  
Db 1847 AAGAAAGAAAGTACTCTCTAATATGCTCTACAAGAAATTCAGAAATGTTGGATGAATTTAC 1906  
Qy 4507 AGTTCCATTCCTCAGTGTGCTGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGG 4566  
Db 1907 AGCTCCCTGAAGGTTTCAGGCTGATTTTCGCTTAAAGATCCCTTATTTCTGAGAGCATAA 1966  
Qy 4567 AGAAATGGTTTATCTCTTTGCCACAACAATTTATAGAAATTTGAAATGAAAGTGGCCACCTG 4626  
Db 1967 AAGAAATGCTTAACGACATTTTGGAACTGCTACCTACAAGTGGGACTTAAAGGTTTCAATCCCA 2026  
Qy 4627 ATGAAGGGATCTCGAGTCCCTCCTGCTGACCTGGAGCAGCTGCGCTTTCACATATCCAG 4686  
Db 2027 ATGAAGGGATCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2086  
Qy 4687 CAATGAAAAATCTATTGGGAGATGAAGAAACCTCTGTTTGGAGCAGCTTCAAAATAGGC 4746  
Db 2087 GCATAGAAGAAATTTTGAGTGAATGAAGATAAACCATTGTTTGGTCTTTTACCTTGCAGAC 2146  
Qy 4747 AGCATATATGCTGAAAGCATTAATGCAAGTTTGCAGTTTGCAGTTCACAGAGGATTAACCTG 4806  
Db 2147 TGGATGACTGCTTAGGTCATTGACGAGATTTTGGCGCAGCACATGACAGTGGCATCAG 2206  
Qy 4807 AGTTCCTGATACAGNAACATCTGGTTCGCTTCTATCAGTTGTTTCTTCTTCAATCAAAAT 4866  
Db 2207 TTTCAAGTGGTCAAGGACATTTTGTAAACCTTTTTCATCACTGCTGCTGCTGCTGCTGCTG 2266  
Qy 4867 CAGAAATACACCACTGCTTCTGCTATAGATCTGTTTTCATGTTTGGTGGTGGTGGTGGTGG 4926  
Db 2267 ATGAGGAATCTCCATGCAATATAGATATGACATGTTTTCATTTATGTTGGTGGCTGGTGG 2326  
Qy 4927 TAGCATTTCCCTCTTGTATTGGGATGACCTGTTTGTATCTGACGCTTCTTTCAGTTAGTT 4986  
Db 2327 TTGCATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2371  
Qy 4987 CTTCCCTATAACCACTTTTATCTCTTCCATTTGATCACCATGCGCACACATGCTTTCAGATAC 5046  
Db 2372 TTGCACTGGAGACCTTTCACATTTTCCATCTGTTTACTATGTCACACATCATACATCT 2431  
Qy 5047 TACTTACAGTAGACAC---AGSCCTACCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5103  
Db 2432 TACTTACTCTATGATACAGAGAGATGCGATGGATCAAGAAATCCCTGCTTGTGAGAAAG 2491  
Qy 5104 CTCATTTCCGATCTTCTTCTTCTGAGAAATTTCTCAATATACAAAGTGGCTCCATTTGGT 5163  
Db 2492 AATCAGCAGTTCTTCTGCTTGTATAAACACTTCCACAGTATACGGAAGTGCCTTGAAG 2551  
Qy 5164 GTGATATTCCTGGCTGTTTGTGGTCTCAGTGAAGAAATGGCAATCACCCCTTATCTTTC 5223  
Db 2552 AAATACCATCCGCTGGCATCTGTGGAGGAGTGTGAGAGTGGAAATCATGCTTCTTCTG 2611  
Qy 5224 GCTG 5283  
Db 2612 AGTCTTCTGCTTATTTTCTTATTTTAAATGGAGTTTCTTCCCTCCCAACCCGACATTTCAAG 2671











XX PF 13-MAY-1999; 99WO-US10602.  
XX PR 14-MAY-1998; 98US-0085426.  
XX PR 15-MAY-1998; 98US-0085537.  
XX PR 15-MAY-1998; 98US-0085696.  
XX PR 21-OCT-1998; 98US-0105234.  
XX PR 27-OCT-1998; 98US-0105877.  
XX (CHIR ) CHIRON CORP.  
XX (HYSE-) HYSEQ INC.  
XX Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;  
XX Reinhard C, Giese K, Randazzo F, Kennedy GC, Pot D, Kassam A;  
XX Lanson G, Drmanac R, Kirkenjakov R, Dickson M, Drmanac S, Labat I;  
XX Leshkowitz D, Kita D, Garcia V, Jones LW, Stache-Crain B;  
XX WPI; 2000-126369/11.  
XX Polynucleotide library used to determine cancerous states of mammalian  
XX cells -  
XX PS Claim 1; Page 916-917; 1097pp; English.  
XX CC AAA00010 to AAA02716 represent polynucleotides isolated from cDNA  
XX libraries constructed from human colon cancer cell lines. The present  
XX invention also describes a method of detecting differentially expressed  
XX genes correlated with a cancerous state of a mammalian cell, comprising  
XX detecting at least one differentially expressed gene product in a test  
XX sample derived from a cell suspected of being cancerous, where detection  
XX of the differentially expressed gene product is correlated with a  
XX cancerous state of the cell from which the test sample was derived.  
XX The polynucleotides sequences can be used in a method for detecting  
XX differentially expressed genes correlated with a cancerous state of a  
XX mammalian cell. The polynucleotides can also be used as probes for  
XX detecting and mapping related genes. They can be used in diagnosis and  
XX prognosis of diseases and disorders (e.g. identification of  
XX pre-metastatic or metastatic cancerous states, stages of cancer, or  
XX responsiveness of cancer to therapy). This is particularly for breast  
XX cancer, oestrogen receptor-positive breast cancer, oestrogen receptor-  
XX negative breast cancer, lung cancer, and colon cancer.  
XX SQ Sequence 756 BP; 219 A; 145 C; 154 G; 219 T; 19 other;  
Query Match 10.1%; Score 638.6; DB 21; Length 756;  
Best Local Similarity 95.9%; Pred. No. 7.2e-164;  
Matches 677; Conservative 0; Mismatches 26; Indels 3; Gaps 3;  
QY 1405 CATATGACCAGCTCATATACAGCCTACAAAGAGCTCTTTGACTGTGAGCTCGCAGAGGCC 1464  
DB 54 CACGAGACCAGCTCATATACAGCCTACAAAGAGCTCTTTGACTGTGAGCTCGCAGAGGCC 113  
QY 1465 AGTTGCATACCACTCCCATTTGACAAAGAGGTCGCGGGCTGTTAAAGCGGAGCTTATG 1524  
DB 114 AGTTGCATACCACTCCCATTTGACAAAGAGGTCGCGGGCTGTTAAAGCGGAGCTTATG 173  
QY 1525 CTGCTTGGCAGGAAGCAAGATATAAAGAGTCATTACAGAAATGCTCTCAACATC 1584  
DB 174 CTGCTTGGCAGGAAGCAAGATATAAAGAGTCATTACAGAAATGCTCTCAACATC 233  
QY 1585 CACTTCATGTAGAGTATTACACTCAGAGTATTGGCTCATCAGAAATTTGCTTGCCTC 1644  
DB 234 CACTTCATGTAGAGTATTACACTCAGAGTATTGGCTCATCAGAAATTTGCTTGCCTC 293  
QY 1645 TTGGTTCCTGGATGAACAAATATTAGCTATTCAAGTGACCTTGGCAGATCTTTTGGC 1704  
DB 294 TNGGTTCTGGATGAACAAATATTAGCTATTCAAGTGACCTTGGCAGATCTTTTGGC 353  
QY 1705 AAGCATGCTTAGAGGAACCTGACTCGGAGAAATCCCTGTCTCATACAGGTTAATGC 1764  
DB 354 AAGCATGCTTAGAGGAACCTGACTCGGAGAAATCCCTGTCTCATACAGGTTAATGC 413  
QY 1765 TTTGGGATGCAAGAGCTTTTATAAAGGTGCCCGCTTATGATCCTTCAATGATCTTCAGCA 1824

DB 414 TTGGGATGCAAGAGCTTTATAAAGGTGCCCGTAAAGTCCCTTCATGAATTCATCTTCAGCA 473  
QY 1825 GTTTTTTATGGAGATGGAATACAAAAAAGCTCTTTGCTATGGAATTTGTGAAGTATTATA 1884  
DB 474 GTTTTTTATGGAGATGGAATACAAAAAAGCTCTTTGCTATGGAATTTGTGAAGTATTATA 533  
QY 1885 AACAACTGCAGAAAGAATATATACAGTATGATCATGACAGAAAGTATCTCTATACTGCAC 1944  
DB 534 AACAACTGCAGAAAGAATATATNAGTATGATCATGACAGAAAGTATCTCTATACTGCAC 593  
QY 1945 TTTCAGTTCAGATGTTTACTGTTCTTACTCTGCTCGACATCTTATTAAGAGCAGAAATG 2004  
DB 594 -TTTCAGTTCAGATGTTTACTGTTCTTACTGTTCTGCTCGACATCTTATTTGAA-AACAGAAATG 651  
QY 2005 TTATCTCTGTCATTACTGAAACCTCTGCTAGAACCTTTTACCTGAGTACTTGGACAGGAACA 2064  
DB 652 TTATCTCTGTCATTACTGAAACCTCTGTTNTAAGTCTTACCTGAGTCTTGGACAGGAACA 711  
QY 2065 ATAAATTCACCTCCAGGGTTATAGCCAGGACAAATTTGGGAAGAGT 2110  
DB 712 NTAAATTCNACTTCCANGGTTAT-GCNGGACANATTGGNAAGATT 756  
RESULT 15  
ABZ70161  
ID ABZ70161 standard; cDNA; 2616 BP.  
XX AC ABZ70161;  
XX DT 24-APR-2003 (first entry)  
XX DE Fibrinogen 9.57 coding sequence.  
XX KW Fibrinogen 9.57; tumour; cytostatic; inflammation; HIV infection;  
XX KW immunological disease; haemopathy; anti-HIV; gene; ss.  
XX OS Unidentified.  
XX FH Key Location/Qualifiers  
XX CDS 415..678  
XX FT /\*tag= a  
XX FT /product= "Fibrinogen 9.57"  
XX CN13611119-A.  
XX PN 31-JUL-2002.  
XX PD 26-DEC-2000; 2000CN-0135939.  
XX PF 26-DEC-2000; 2000CN-0135939.  
XX PR (BODE-) BODE GENE DEV CO LTD SHANGHAI.  
XX PI Mao Y, Xie Y;  
XX WPI; 2002-751538/82.  
XX P-PSDB; ABF59141.  
XX New polypeptide fibrinogen 9.57 and polynucleotides encoding this  
XX polypeptide -  
XX PS Claim 6; Page 24-25 (Disclosure); 33pp; Chinese.  
XX CC The present sequence is the coding sequence for fibrinogen 9.57. The  
XX protein is useful for treating various diseases, such as malignant  
XX tumours, inflammations, immunological diseases, haemopathy and HIV  
XX infection.  
XX SQ Sequence 2616 BP; 780 A; 449 C; 512 G; 875 T; 0 other;  
Query Match 9.9%; Score 626.8; DB 24; Length 2616;  
Best Local Similarity 99.7%; Pred. No. 2.6e-160;

		Matches	628;	Conservative	0;	Mismatches	2;	Indels	0;	Gaps	0;
QY	5679	GCCGGAGTCTGCATTTTCATAAAATCAGAGATGCCGAGTGTCTGTGTTGAAGTAAA									
Db	1										
QY	5739	GCCAGAGGCTGTGCCCTATCCAGCTCTTACTTGGATGAATATGAGAAACAGACCCCTGGC									
Db	61										
QY	5799	CTGAAGAGGGCAACCCCTTCATTTATCTCGTGAGCGGTATCGGAAGTCCCATTTGGTC									
Db	121										
QY	5859	TGGCAACACACTGCATTATAGAAGAGATTGCTAGGAGCCAGAGACTAATCAGATGTTA									
Db	181										
QY	5919	TTTGGATTCAACTGGCAGTTACTGTGAGCTCCAACCTCGCTCAAGACAATCACAAATGA									
Db	241										
QY	5979	CGACAGTACTAAGCTGATTCAAAATTTATGGAATTTCTCAGGGCTGGGAAGTATT									
Db	301										
QY	6039	GGAGGTCCTTTGCTCCATGTCCAGGTTCACTTACATCAATAAATATTCTTAAATGGAG									
Db	361										
QY	6099	TATTGCTTCAATTAGCAACATATGCTTCACAGGAAAAAGGACATAGATCAATCTGTT									
Db	421										
QY	6159	TTATGCTAGTATTTCAGGAATTATTCCCTTCATAATTGTCTCATTTTAT									
Db	481										
QY	6219	TTCATCCACTTGGTAGATCAAGTCAGTCAACAGTTGTAGACATTTTATGTGTGGTTA									
Db	541										
QY	6279	ACTCTTCTGCAATTTTGTATTTGGTGT									
Db	601										

Search completed: September 27, 2003, 01:57:39  
Job time : 1013.89 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 27, 2003, 01:01:23 ; Search time 267.924 Seconds  
(without alignments)  
10391.910 Million cell updates/sec

Title: US-09-724-126A-1  
Perfect score: 6308  
Sequence: 1 gccagaattgcacgaggg.....aattttgtatttggtgtttt 6308

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA:\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4435.4	70.3	6395	2	US-08-982-956-1
2	4435.4	70.3	6395	3	US-09-228-317-1
3	996.2	15.8	1001	2	US-08-982-956-2
4	996.2	15.8	1001	3	US-09-228-317-2
5	57.4	0.9	7218	1	US-08-232-463-14
6	52.2	0.8	7218	1	US-08-232-463-14
7	45.6	0.7	111282	4	US-09-754-250-3
8	44.4	0.7	7333	3	US-09-138-024-21
9	44.4	0.7	7333	4	US-09-404-066-21
10	44.4	0.7	7333	4	US-09-573-322-21
11	44.2	0.7	72928	3	US-09-009-913-1
12	42.8	0.7	50000	4	US-09-146-053-4
13	42.2	0.7	325	3	US-08-991-789A-236
14	42.2	0.7	325	4	US-09-062-451-236
15	42.2	0.7	325	4	US-09-598-326-236
16	42.2	0.7	325	4	US-09-289-198-236
17	41.2	0.7	36551	4	US-09-738-894A-3
18	41.2	0.7	36551	4	US-09-964-469-3
19	41.2	0.7	42571	4	US-09-810-347-3
20	41	0.6	1517	3	US-08-963-602-6
21	40.6	0.6	41708	4	US-09-470-512A-3
22	40.4	0.6	152331	3	US-09-128-155-16
23	40.2	0.6	9310	4	US-09-453-702B-256
24	39.8	0.6	2861	1	US-08-299-953-1
25	39.8	0.6	2861	1	US-08-459-415-1
26	39.8	0.6	2861	4	US-09-066-687-1
27	39.8	0.6	2861	5	PCT-US95-11231-1

c 28	39.8	0.6	3881	1	US-08-299-953-2	Sequence 2, Appli
c 29	39.8	0.6	3881	1	US-08-459-415-2	Sequence 2, Appli
c 30	39.8	0.6	3881	4	US-09-066-687-2	Sequence 2, Appli
c 31	39.8	0.6	3881	5	PCT-US95-11231-2	Sequence 2, Appli
c 32	39.8	0.6	4376	1	US-08-119-125A-1	Sequence 1, Appli
c 33	39.4	0.6	751	4	US-09-687-698-11	Sequence 11, Appli
c 34	39.2	0.6	64467	4	US-09-803-671B-3	Sequence 3, Appli
c 35	39	0.6	32042	4	US-09-245-281-44	Sequence 44, Appli
c 36	39	0.6	32042	4	US-09-340-620A-63	Sequence 63, Appli
c 37	38.8	0.6	246240	2	US-08-724-394A-20	Sequence 20, Appli
c 38	38.8	0.6	246240	2	US-08-724-394A-22	Sequence 21, Appli
c 39	38.8	0.6	246240	2	US-08-724-394A-22	Sequence 22, Appli
c 40	38.4	0.6	5703	4	US-09-280-590A-36	Sequence 36, Appli
c 41	38.4	0.6	18596	3	US-09-318-448-11	Sequence 11, Appli
c 42	38.4	0.6	18597	4	US-09-962-665-8	Sequence 8, Appli
c 43	38	0.6	1785	4	US-09-601-198-156	Sequence 156, App
c 44	37.8	0.6	19011	1	US-08-310-356-36	Sequence 36, Appli
c 45	37.8	0.6	19557	5	PCT-US92-06300-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1  
US-08-982-956-1  
; Sequence 1, Application US/08982956  
; Patent No. 5861312  
; GENERAL INFORMATION:  
; APPLICANT: Varshavsky, Alexander  
; APPLICANT: Kwon, Yong Tae  
; TITLE OF INVENTION: NUCLEIC ACID ENCODING MAMMALIAN UBRI  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Kevin M. Farrell, P.C.  
; STREET: P.O. Box 999  
; CITY: York Harbor  
; STATE: ME  
; COUNTRY: US  
; ZIP: 03911  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/982,956  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Farrell, Kevin M.  
; REGISTRATION NUMBER: 35,505  
; REFERENCE/DOCKET NUMBER: CIT-2001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (207) 363-0558  
; TELEFAX: (207) 363-0528  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 6395 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 115..5385  
US-08-982-956-1

Query Match 70.3% Score 4435.4; DB 2; Length 6395;  
Best Local Similarity 86.3%; Pred. No. 0;  
Matches 4997; Conservative 0; Mismatches 721; Indels 69; Gaps 6;  
Qy 590 TTCAGGGCGCTCGTAAAGTGTGCTGCTCTC-TCCGACCGCCACAGTTTCGCT 648  
|||||







4388 CCGTGGATCTGCAGCGCTGCCACTTAGTTCTTTCATATAACCACCTCTATCTCTTCCATC 4447  
5017 TGATCACCATTGGCACACATGCTTCAGATACACTTACAGTAGACA-----CAG 5064  
4448 TGATCACCATTGGGCACATGCTTCAGATACACTTTCATACAGATACAGATCTGTCTCCAG 4507  
5065 GCCTACCCCTTGTCTCAGGTTCAAGACAGACAGTGAAGAGGTCATTCGCCATCTCTTCTTCT 5124  
4508 GCGCGCGCTGTGTAGGGTGGAAGAGATAGTGAAGAGGTCGCTGTGCATCTCTCTTCT 4567  
5125 TTGCAGAAATTTCTCAATATACAAAGTGGCTCCATTGGTGTGTGATTTCTTGGCTGGGTATT 5184  
4568 TTGTGGAAGTGTCCGACACACAGACGGCTCTACTGGGTGGGTGCTCCCGGCTGGTACC 4627  
5185 TGTGGGFTCTCACTGAAGAATGGCATCACCCCTTATCTTCGCTGTGCTGCATTTGTTTTCC 5244  
4628 TGTGGCTCTCCCTGAGGAACGGCATCACCCCTTACCTCCGCTGTGCTGCACCTGCTTTTCC 4687  
5245 ACTATTTTACTTGGGTAACTCCGCTGAGGAACCTGCATACAAATTCCTGCAAGAGGAGT 5304  
4688 ACTATTTTACTTGGGTAACTCCGCTGAGGAACCTGTTTGGCAATTCCTGCTGAAGGAGAT 4747  
5305 ACAGTGCACCTGTAGCTATCTATCTTACCTACAAATTTGTTCCCTCTCTTCCAGGAAT 5364  
4748 TCAGTGCACCTGTAGCTATCTATCTTACCCACAAATTTGTTCCCTCTCTTCCAGGAAT 4807  
5365 ATTGGGATCTGAAGCCCTTGTCCAGAGGTGGTGTGAGATCTTCCCTTACTAACT 5424  
4808 ATTGGGATCTGAAGCCCTTACTACAGAGGTGGTGTGAGATCTTCCCTTACTCAAGT 4867  
5425 GTTTGAAGCAAAAACACCGTGTGAGTACCTTAGAAAAGAAATAGTTTGTATAGAGC 5484  
4868 CTTTGAAGCAAAAAGTGTGTGCTAGTACCTTAGAAAAGAAATAGTTTGTATAGAGC 4927  
5485 TTCTGTAGTACTATAGTGTCTCTGAATCAAGTCTCTATTCAGGTGCCACGGTCTG 5544  
4928 TTCTGTAGGACTACAGTGTCTTCTAAATCAGGCTTCTCACTTAGGTGTCCACGGTCTG 4987  
5545 CAGATGATGAGCAAGCATCTGTCTCTGCTCTGCTTTCTGTGGGCTATACATATGTTCTC 5604  
4988 CAGATGATGAGCAAGCATCTGTCTCTGCTTTCTGTGGGCTATCTGTGTTCTC 5047  
5605 AGAACAATTTGCTGCCAGGAATTTGTGAACGGGGAAGAGTTGGAGCTTGCATTTTTCACG 5664  
5048 AGAACAATCTTGTGCAAGAAATAGTGAATGGGAAGAGTTGGAGCGTGTGCTTTTTCATG 5107  
5665 CACTTCACTGTGGCCGGAGTGTGCATTTTCTTAAATCAGAGATGCCAGTGGTCC 5724  
5108 CGCTTCACTGTGGTGTGGAGTGTGCATTTTCTTAAATCCGAGAAATGCAGGTTGGTCC 5167  
5725 TGGTTGAAGTAAAGCCAGAGGCTGTGCCTTATCCAGCTCTCTTACTTGGATGAATATGAG 5784  
5168 TGGTGAAGAAAGCCAGAGGCTGTGCCTTATCCAGCCCTTACTTGGATGAATATGAG 5227  
5785 AAACAGACCCCTGCTGAAGAGGGGCAACCCCTTCAATTTATCTGTCGAGCGGTATCGGA 5844  
5228 AAACAGACCCAGGCTAAAGAGAGGAAACCCCTTCAATTTATCTGCGGAGCGGTATCGGA 5287  
5845 AGCTCCATTTGGTGTGCAACACACATGCAATATAGAAGAGATGTCTAGAGCCCAAGAGA 5904  
5288 AGCTGCATTTGGTGTGCAACACACATGCAATATAGAAGAGATGTCTAGAGCCCAAGAGA 5347  
5905 CTAATCAGATGTTATTTGGATTTCAACTGGCAGTGTACTGTGAGTCCCAACTCTGCCTCAAG 5964  
5348 CTAATCAGATGCTATTTGGATTTTAACTGGCAGTGTACTGTGAGTCTTCTGAGTCTCAAG 5407  
5965 ACAATCACAATAGCAGCAGTAGTAAAGGCTGATTTCAAAATATATGGAATACTTTCTGAGG 6024  
5408 ACAATCATGATGACATCAATAAAGACTGATCTAAATTTCTAGAGAACTTTCTGAGG 5467  
6025 GCTGGGAAAGTATGGAGGCTTTTGTCTCCATGTCCAGGTTCACTTACATCAATAAAT 6084  
5468 ACGGGGGAAGTATGGAGGCTTTTGTATCCATGTCTCCAGATTTACACACATTAATAAAT 5527

QY 6085 ATTCTTAAATGGAGTATGCTTTTCAATTAGCAAAACATATGCTTCACAGGAAAAA-AGGAC 6143  
DB 5528 ATTCTTAAATGGAGTATGCTTTTCAATTAGCAAAACATATGCTTCACAGGAAAAA-AGGAC 5587  
QY 6144 ATAGATCAATCTGT-----TCTT 6160  
DB 5588 ATAGATCAATCTGT-----TCTT 5647  
QY 6161 ATGCTAGTATTTCCAGGAATTTATGCCCTTCAATAATTTGCTCATTCTTCTTATTT 6220  
DB 5648 TTTGCTAATTTTCCAGAAACTCTTCCCTTCACTAATCTCTCTAGTTCATTTCAATAT 5707  
QY 6221 CATCCACTTGTAGTAGAAGTACGTCACAAAGTGTGTAGACATTTTATGTTGGTTAA 6280  
DB 5708 CACCACCTGTTAATGAGTACATTAAGCATTTTGTGGACATTTCTCCATCTGGCTAAC 5767  
QY 6281 TCTTCTGCAATTTTGTATTTGGTGT 6307  
DB 5768 ATCTGTGCACCTTTGTATTTGGTGT 5794

RESULT 2  
US-09-228-317-1  
; Sequence 1, Application US/09228317  
; Patent No. 6159732  
; GENERAL INFORMATION:  
; APPLICANT: Varshavsky, Alexander  
; APPLICANT: Kwon, Yong Tae  
; TITLE OF INVENTION: NUCLEIC ACID ENCODING MAMMALIAN UBRI  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Kevin M. Farrell, P.C.  
; STREET: P.O. Box 999  
; CITY: York Harbor  
; STATE: ME  
; COUNTRY: US  
; ZIP: 03911  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/228,317  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Farrell, Kevin M.  
; REGISTRATION NUMBER: 35,505  
; REFERENCE/DOCKET NUMBER: CIT-2001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (207) 363-0558  
; TELEFAX: (207) 363-0528  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 6395 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 115..5385  
US-09-228-317-1

Query Match 70.3%; Score 4435.4; DB 3; Length 6395;  
Best Local Similarity 86.3%; Pred. No. 0;  
Matches 4997; Conservative 0; Mismatches 721; Indels 69; Gaps 6;  
QY 590 TTCAGGGGCGCTGCTGTAAGATGCTGCTCTGCTC-TCCGACCGGCCACAGGTTTCCGCT 648  
DB 8 TTCAGGGGCGCTGCTGTAAGATGCTGCTCTGCTGCGGCTCGGCGGCCACAGGTTTCCGCT 67

QY	649	TGCTCTGCGCGGGTTCGCAACTGCAGCGTCAGTTCCTCCCAAGATGGCGGAGG	708
Db	68	AGCTGGCGCGGGGTGCGGAACCTCGGGCGTTCGTTCCCTTAAGATGGCGGAGG	127
QY	709	AGGCTGGAGTACTGAGAGGATGGAATCAGCGCGAGTTACCCAGACCCCTCAGCGTC	768
Db	128	AGATGACGCGCGGAGAGGATGGAGCTCAGCCCGAGCCTCCCTGGCCCGCAGCGGC	187
QY	769	TGGCATCTTGGTGGGATCAGCAAGTTGATTTTATACTGCTTCTTTCATCATTTGGCAC	828
Db	188	CGGCATCGTGGTGGGATCAGCAAGTTGATTTCTATCTGCTTCTTACATCAATTTGSCAC	247
QY	829	AATTGGTCCAGAAATTTACTTTCCTGTAATGGACCCAGACTTGGAAAAGCAGAGAAA	888
Db	248	AATTAGTCCAGAAATTTTTCCTGAGATGGACCCAGATTGGAAAAGCAGAGAGA	307
QY	889	GTGTCAAAATGCAATATTACTCCACTCGAATGGTACTTATTTGAGAGAGATCCAGATA	948
Db	308	GTGTACAGATGCAATACTCACTCCTTTGGAGTGGTACTTATTTGAGAGGATCCGATA	367
QY	949	TTTGCTTAGAGAAATTTAGACACAGTGGAGCATTTTCAGCTTTTGGGAGGTTTCAAAA	1008
Db	368	TTTGCTTAGAGAAATTTAAACACAGTGGAGCTTCCAGTTGTGTGGGAGGTTTCAAAA	427
QY	1009	GTGGAGACAACTTATCTTCGAGGATTTGCAATTTGATCCAACTGTGTACTCTGTA	1068
Db	428	GTGGAGAACACATATTTCTGTAGGATTTGCAATTTGATCCAACTGTGTACTCTGTA	487
QY	1069	TGGACTGCTTCAGGACAGTGTTCATAAAATCATCGTTACAGATGCAATCTTACTG	1128
Db	488	TGGACTGCTTCCAAAGTAGTTTCATAAAACCCTCGTTACAAGATGCAATCTTACTG	547
QY	1129	GAGGAGGTTCTGTGACTGTGGAGACACAGAGGCTGGAACCTGGCCCTTTTGTGTAA	1188
Db	548	GAGGAGGTTCTGTGACTGTGGAGACACAGAGGCTGGAACCTGGCCCTTTTGTGTG	607
QY	1189	ATCATGAACCTGGAAGACAGGTACTATATAAGAGAAATTCACGCTGTCCGTTGAAATGA	1248
Db	608	ATCAGAGCCTGGAAGACAGGTACTATCAAAAGAGAGCTTACATTTGCCATTTCAATGA	667
QY	1249	AGGTAAATTCGAAGCAGGAAATTTTCTTCAGTGATATAATATGCTGATAGATGA	1308
Db	668	AGGTAAATTCGAAGCAGGAAATTTTCTTCAGTGATATAATATGCTGATAGATGA	727
QY	1309	CTATATGGAAGAGGAAAGAACTGCCCTCCTGAACCTCCAGATAAGGGAGAAAATGAAA	1368
Db	728	CTATATGGAAGAGGAAAGAAATTCGCCCTCCTGAACCTGCAGATTAAGGGAGAAAATGA	787
QY	1369	GATACATTTGTCTCTTTCAACGATGAGCACCATTTCGTATGATGATGATGATGATGATG	1428
Db	788	GATACATTTGTCTCTTTCAACGATGAGCACCATTTCGTATGATGATGATGATGATGATG	847
QY	1429	TACAAGAGCTCTTGACTGTAGCTCGAGAGGCCAGTTGCAATACCACTGCCATTTGACA	1488
Db	848	TGCAGAGAGCTCTAGATTTGGAGCTTGCAGAGGACAGCTGCCACGAGCTGCCATGACA	907
QY	1489	AGAGGGTCTGGGGTGTAAAGCGGAGCTTATCTGCTGCCAGGAGCAAGGAAG	1548
Db	908	AGAGGGTCTGGGGTGTAAAGCGGAGCTTATCTGCTGCCAGGAGCAAGGAAG	967
QY	1549	ATATAAGAGTCAATTCAGAAAATGCTCTCAACATCCACTTCATGTAGAGATTTACAT	1608
Db	968	ATATAAGAGTCACTCAGAGAGCTCTCTCAGGACCCCTCCATGTGGAAGTCTGCCT	1027
QY	1609	CAGAGATTATGCTCATCAAAATTTGCTTTGGCTTGTGGTCTTCTGGATGAACAAATTA	1668
Db	1028	CCGTGGTTATGCTCACCAGAAATTCGCTCTGGCCCTTGGCTTCTTGGATGAACAAATTA	1087
QY	1669	TGAGCTATTCAAGTGTACTTAGCAGATCTTTTGGCAAGCATGCTTTAGAGAGACCTG	1728
Db	1088	TGAGCTATTCAAGTGTACTTAGCAGATCTTTTGGCAAGCATGCTTTAGAGAGACCTG	1147
QY	1729	ACTCGGAGAAATCCCTGTCTCATAGCAGGTTAATGCTTTGGATGCCAAGCTTTTAAAG	1788
Db	1148	GCTCTGAAAATCCCTGTCTTATAAGCAGACTAAATGCTTTGGATGCCAAGCTTTTAAAG	1207
QY	1789	GTGCCCCAAGATCCCTCATGAATTTGATCTTTCAGCAGTCTTTTATGAGATGGAATACA	1848
Db	1208	GTGCCCCAAGATCCCTCATGAATTTGATCTTTCAGTCTTTTATGAGATGGAATACA	1267
QY	1849	AAAACTCTTTGCTATGGAATTTGAGATTTATAAACAACCTGCAGAAAAGATATATCA	1908
Db	1268	AAAACTCTTTGCTATGGAATTTGAGATTTATAAACAACCTGCAGAAAAGATATATCA	1327
QY	1909	GTGATGATCATGACAGAGATCTCTAATCTGACCTTTCAGTCTGATGTTTACTGTTTC	1968
Db	1328	GCGAGCACCAGAGAGAGCATCTCCATAACCGCCCTCTCGTGAGATGCTCACCGTCC	1387
QY	1969	CTACTCTGGCTCGACATCTTATTTGAAGCAGAGATTTTATCTCTCTCTATTTACTGAACTC	2028
Db	1388	CGACTTTGGCCCGCATCTTATTTGAAGCAGAGATTTTATTTCTCTCATTTACTGAAACGC	1447
QY	2029	TGCTAGAAGTTTACCTGAGTACTTTGGACAGGACAAATAAATTTCAACTTCCAGGGTTATA	2088
Db	1448	TGCTAGAAGTTTACCTGAGTACTTTGGACAGGACAAATAAATTTCAACTTCCAGGGTTATA	1507
QY	2089	GCCAGGACAAATTCGGGAAGATATATGCAGTAATATGTGACCTAAAGTATATCTCGATCA	2148
Db	1508	GCCAGGACAAATTCGGGAAGATATATGCAGTAATATGTGACCTAAAGTATATCTCGATCA	1567
QY	2149	GCAAAACCACAATATGGACAGAAATTAAGAAATGACGTTCTTCAAGGTTTTCGATCTT	2208
Db	1568	GCAAGCCTGTATATGACAGACAGATTTAAGAGCGCAGTTCTTGAAGGTTTCCGGTCTT	1627
QY	2209	TTTTGAAGATCTTTACCTGTATGCGGGAATGGAAGAAATCCGAAGACAGCTTTGGCGAAC	2268
Db	1628	TTCTGAAGATCTTTACCTGTATGCGGGAATGGAAGAAATCAGAAGACAAAGTTGGACAAC	1687
QY	2269	ACATTGAAGTGGATCTTGGAGGCTGCCATTTGCTATACAGATGCAATTTGAAGAATA	2328
Db	1688	ACATTGAAGTGGAGCTTGGAGGCTGCCATTTGCTATACAGATGCAATTTGAAGAATA	1747
QY	2329	TTTTTACCTCTTCCAGAGTGTGCTCTGTGATGAAGAACTCTTACTTGTGGCTTATA	2388
Db	1748	TTTTTGTCTATGTTCCAGAGTGGTGTGCTTGTGATGAAGATCTCTTACTTGTGGCTTATA	1807
QY	2389	AAGAATGTCACAAAGCTGTGATGAGGTGCAGTACCAGTTTTCATATCTAGTAGCAAGACAG	2448
Db	1808	AAGAATGTCACAAAGCTGTGATGAGGTGCAGTACAAATTTTCATGTCAGTACCAGACAG	1867
QY	2449	TAGTACAACTCGTGGACATAGTTTGGAAACAAAGTCTACAGATATCTGAGGATCTTG	2508
Db	1868	TAGTGAATTTGTGGGTCATAGTCTGGAACCAATCTTACAAAGTCTGAGGAGCTTG	1927
QY	2509	TAAGCATACATCTGCCACTCTTAGGACCTTGTGGTCTTCATCTAGTCTTAAACGAGC	2568
Db	1928	TAAGCATACATCTGCCACTCTTAGAACACTTGTGGTCTTCATCTAGTCTTAAACGAGC	1987
QY	2569	TGGTGTCTGTTTCAAGACTGCAATTTGTCTTTTGGAGACTTTTCAAGTAGAGGTAC	2628
Db	1988	TAGTGTCTATTTCAAGACTGCAATTTGTGCTTTTGGAGACTTTTCAAGTAGAGGTCC	2047
QY	2629	TAGTGAATATCTTTTACGTTGTGTTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGT	2688
Db	2048	TGGTGGAGTACCCCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	2107
QY	2689	GAGAAATGAGCTGTCTTATTTAGCCAGGTGTTTTTATACAGATGTTTAAAGTCCAG	2748
Db	2108	GAGAAATGAGCTGTCTTATTTAGCCAGGTGTTTTTATACAGATGTTTAAAGTCCAG	2167
QY	2749	AAGAAATGATGATAAAGATATCATGCTTTCAGATTTGCTGCTTATTTGATGCCA	2808
Db	2168	AGAAATGATGATAAAGATATCATGCTTTCAGATTTGCTGCTTATTTGATGCCA	2227
QY	2809	ATAAGTTCTTCTTACTTGTGATGAGGTATGAATTTGCCGAGGCTTTTAAACAAGACCA	2868

Db	2228	ACAGTCTTGTGTTACTTGGTACTTCAGAGATATGACTTACTGATGCTTTTAAACAAGACCA	2287	Db	3308	AAGACTCCATTATGGAGGAAGAGAGCACCTCAGCAGTCAGTCAGGCCCTCTAGAATTGCTC	3367
Qy	2869	TATCTACAAAGACCAAGGATTTGATTAAACAATATAATACACATAATAAGAAATGCTTC	2928	Qy	3940	TGGGTCTCTAAACGGGGTCCATCTGTACTGAAAAAGAGGTGCTGACGTGCATCTTTTGCC	3999
Db	2288	TATCCACAAAGACCAAGGATTTGATTAAACAGATATAATACATTAATAGAAATGCTTC	2347	Db	3368	TGGGCCCTTAAACGGGGCCGGCTGTACCGAAAAGGAGGTGCTGACGTGCATCTCTGCC	3427
Qy	2929	AGTCTCTCATATATGTTGGGTGAGGTTATGTACTCTGAGTGGGAAATGTGACCAAG	2988	Qy	4000	AAGAARACAGAGGTTGAAATAGAAAATAATGCCATGTTATATCGCCCTCTGTCCAGA	4059
Db	2348	AGGCTCTCATCTATATGTTGGGAACGTTATGTACTCTGAGTGGGAAATGTTACACAG	2407	Db	3428	AAGAAGAACAAAGAGGTAAACCTAGAAAATAATGCCATGTTATGTACAGCATGTGTCAGA	3487
Qy	2989	AACAGTTCACAATGAGAGAAATCATTCACCTTGCTTTGCAATGAACCCATGCCACACAGT	3048	Qy	4060	AATCTACTGCTTAACCCAGCACAGGGGAAAACCCATAGAACTCTCAGGAGAACCCCTAG	4119
Db	2408	AGGAGGTTAATAGAGAGATTTACTCATTCTTTGCAATGAGCCCATGCCACACAGT	2467	Db	3488	AATCCACGCCCTTACCAGCACAGAGGGAAGCCCTGTGGACCACTTAGGGGAAACACTGG	3547
Qy	3049	CCATTGCCAAAAATTTACTTGAGAATCAAAATAATGAACTGGCTTAGAGAAATGTCATAA	3108	Qy	4120	ACCACCTTTTCATGATCCAGACTTTGGCATATGGAACCTATACAGAAAGCTGTGGTCATG	4179
Db	2468	CCATGCCGAAACCTTACCTGAGACAGCAAAATATGNAACCTGGCTTAGAGAAATGTCATAA	2527	Db	3548	ACCCTCTTTTCATGATCCAGACTTTGGCACATGGAACCTTATACAGAAAGCTGTGGTCATG	3607
Qy	3109	ACAAAGTGGCCACATTTAAGAAACCCAGGTGTATCAGGCCCATGGAGTTTATGAACATAAG	3168	Qy	4180	TAATGCACGAGTGTGCTGGCAGAAATATTTTGAAGCTGTACAGCTGAGCTCTCAGCAGC	4239
Db	2528	ACAAAGTGGCCACATTTAAGAAACCCAGGTGTGTCGGGCCATGGAGTTTATGAATTTGAAG	2587	Db	3608	TAATGSCATGCACTGTGCTGGCAGAAATATTTTGAAGCTGTGACGCTGAGCTCCGACAGC	3667
Qy	3169	ATGAATCAGTGAAGACTTCAATATGTACTTTTATCATTTACTCCAAACCCAGCATAGCA	3228	Qy	4240	GCATTCTGTTGACCTTTTGGACTTTGAAAGTGGAGAATATCTTTGCCCTCTTTGCAAAAT	4299
Db	2588	ATGAATCAGTGAAGACTTCAATATGTACTTTTACCAATTTATCTAAACACACAGCATAGCA	2647	Db	3668	GCATTCTGAGACTGTGTTGACCTGGAGAGCGGCGAGTACCTATGCCGCTCTGCAAGT	3727
Qy	3229	AGGCTGAACATATGCAGAGAAAAGGAGAAAACAAAGATGAAGACTTTGCGGC	3288	Qy	4300	CTCTGTGCAATPACTGTGATCCCAATTTATCTTTGCAACCTCAAAAGATAAACAAGTGA	4359
Db	2648	AGGCTGAACATATGCAGAGAAAAGGAGAAAACAAAGATGAAGACTTTGCGGC	2707	Db	3728	CTCTCTGCAACACTGTCTATCCCATCATCTTTGCAGCCGAGAAGATCAACAGTGA	3787
Qy	3289	CACACACACTCTGAAATCTGCCCTGTCTTACAGCAAAAGTGAATTAACCTTCTCAACTGTG	3348	Qy	4360	ATGCAGATGCTTCTCAACTTTTGACCTGGCAGCGGTGGATACAGACTGTTCTGGCCA	4419
Db	2708	CGCCACCTCTCCAGAGTTCTGCCCTGTCTTACAGCAAAAGTAGTCAACCTGTCTCAGCTGTG	2767	Db	3788	ATGGGAGGCTCTTGTCTCAACTTTTGACCTTTGGCCGGTGGATACAGACTGTCTTGCCA	3847
Qy	3349	ATATCATGATGTACATCTCAGGACCCGATTTTGAGCGGGCAATAGACACAGATCTTAAC	3408	Qy	4420	GAATATCAGGTTATAATATAAGACATGCTAAAGGAGAAAAAC--CAATTCCTATTTTCT	4476
Db	2768	ATGTTATGATATACATCCTCAGGACCCATCTTGAGCGGGCAATAGACACAGGAGTCTAATC	2827	Db	3848	GAATATCGGGTTATTAATATAAAGCAATGCTAAAGGAGAGAGCCCGGAGTCTCTGTCTGT	3907
Qy	3409	TGTGGACCGAAGGATGCTCCAAATGGCTTTTCATATCTTGGCAATTTGGTTTACTAGAAG	3468	Qy	4477	TTAATCAAGGAATGGGAGATTTCTACTTTGGAGTTCCCATCTCCATCTCTGAGTTTGGGTTG	4536
Db	2828	TGTGGACAGAGGATGCTCGAGATGGCTTCCATATATTTGGCACTGGGCTTGTGGAAG	2887	Db	3908	TTAATCAAGGAATGGGAGTTCAACTTTTGTAGCTTTTCAATTCATCTCTGAGTTTGGAGTTC	3967
Qy	3469	AGAAGCAACAGCTTCAAAAAGCTCTCGAAGAAAGTGAATTAACCTTTATCATAAAG	3528	Qy	4537	AGCTCTCGATTAATATTAATCAATACATCAAGAAATAGGTTATTTCTTTGGCACAACAA	4596
Db	2888	AGAAGCAGCAGCTTCAAAAAGCTCTCGAAGAAAGTGGCTTTTGAATTTTACCATAAG	2947	Db	3968	AGCTCTCGGTTGAAATTTCAATATAGTATCAAGAAATGGTCTATTTCTTCGCGCACAA	4027
Qy	3529	CTTCAAGATTGGGAAGTTCAAGCTT-----GAATATACAATGCTTTTGGAAAAAC	3579	Qy	4597	TTTATAGAAATTGGAATTGAAAGTGCCACCTGTATGAAGGGATCCTCGAGTCCCATCTCA	4656
Db	2948	CTTCAAGATTGGGAAGTTCAAGCTTCAAGATATCAAAATGCTCTTGGAAAGAC	3007	Db	4028	TTTACAGAAATTTGGCCTGAAAGTGCCCTCTGTATGAACCTAGACCCAGAGTGGCCATGATGA	4087
Qy	3580	TCAAAGGAATTCCTCAGCTTGAAGGCCAGAGGACATGATAACGTGGATACCTTCAGATGT	3639	Qy	4657	CCTGGAGCACCTTGCCTTTTCACTATCCAGGCAATTTGAAAATCTATTGGGAGATGAAGGAA	4716
Db	3008	TCAAAGGAATTCCTCAGCTTGAAGGCCAGAGGACATGATAACGTGGATCTCCAGATGT	3067	Db	4088	CCTGGAGCACCTTGCCTTTTCACTATCCAGGCAATTTGAAAATCTATTGGGAGATGAAGGAA	4147
Qy	3640	TTGACACAGTGAAGCGATTGAAGAAAAATCTTGTAAATTTAGTGAACCAACATCAGAT	3699	Qy	4717	AACCTCTCTTTGGAGCAGCTTCAAAATAGCAGCATTAATGGTCTGAAGCATTAATCGAT	4776
Db	3068	TTGACACAGTGAAGCGATTGAAGAAAAATCTTGTAGTTGTGCAACCACTTCAGGAC	3127	Db	4148	AACCTCTATTTGGAGCAGCTTCAAAATAGACGATAGCGGTCTGAAGGCGCTTAATCGAT	4207
Qy	3700	CGGAATCTTAAAGAAATGATGAGATTACTCATGATTAAGAAAAAGCAGCAAGAAAGAA	3759	Qy	4777	TTGCAAGTTCACAGAGGATTAACCTGCTCAGCTCTGTATACAGAAACATCTGGTTCCGTC	4836
Db	3128	TGGAGTGCATTAAGAGTGAAGAGATTACTCATGATTAAGAAAAAGCAGCAAGAAAGAA	3187	Db	4208	TTGCAAGTTCACAGAGGCTACCTGCCCTCAGGTTCTGTATACAAACATCTGGCTCGGC	4267
Qy	3760	AAGCTGAGCTGTAGGCTACATCGCCAGAGATCATGGCTCAGATGCTCGCTTACAGAT	3819	Qy	4837	TTCTATCAGTTTCTTCTCTTAACATAAATCAGAAGATACACATCCCTTCTGTCTATAG	4896
Db	3188	AAGCTGAGGCGGCTTAGGCTTCATCGCCAGAGATCATGGGCCAGATGTCTGCCCTTACAGA	3247	Db	4268	TCCTGTCTAGTTATTTCTTAACCTGCAATCAGAAATACACAGGCCCTTCTGTCTGTGG	4327
Qy	3820	AAAACCTTCAATGAAACTCATAACTCATGTATGACAAATACATCAGAAATGCTGGGAAAG	3879	Qy	4837	ATCTCTTTTCATGTTTGGTGGTGTGTGTAGCATTCACATTCCTTGTATGGGATGACC	4956
Db	3248	AAAACCTTCAATGAAACCCACAACTCATGTATGATATACTCAGAGTACACAGGAAG	3307	Db	4328	ATCTCTTCCATGTTCTGTGGTGGCGAGTCTTAGCGTTCCCATCTTGTATTTGGGATGACA	4387
Qy	3880	AAGATTCATATGAGGAAGAGAGACCCAGCAGTCAGTGAATCTTCTAGAAATGCTT	3939	Qy	4957	CTGTTGATCTGACGCTTCTTCTAGTTAGTTCTTCTCTATTAACCACTTTTATCTTCCATT	5016
				Db	4388	CCGTGGATCTGCGCGCTGCCACTTAGTTCTTCTATATAACCACTCTATCTCTTCCATC	4447

QY		5017	TGATCACCATGGGCACACATGCTTCAGATACTACTTACAGTAGACA-----CAG	5064
Db		4448		4507
QY		5065	GCCATTACCCTTGCTCAGGTCCAAGAAGACAGTGAAGAGGCTCAATCCGCATCTCTTTCT	5124
Db		4508		4567
QY		5125	TTGCAGAAATTTCTCAATATACAAAGTGGCTCCATPTGGGTGTGATPATTCCTGGCTGGPATT	5184
Db		4568	TTTTGGAAGTGTGCGACACACAGACGCCCTCACTGGGTGCGGTGCTCCGGCTGGTAAC	4627
QY		5185	TGTGGGTCTCATTGAAGAAATGGCATCACCCCCTTATCTTCCTGTGTGCAATGTTTTC	5244
Db		4628	TGTGGCTCTCCCTTGAGGAACGGCATCCCCCTTACCTCCGCTGTGTGTGCACTGCTTTTC	4687
QY		5245	ACTATTACTTTGGGTAAACTCCGCTTGAGAACTGCATACCAATTTCTGCAGAAGGAGAT	5304
Db		4688	ACATTTTACTTTGGAGTAGCTCCGCGCTGAAGAACTGTTGGCAAATCTGCTGAAGGAGAAT	4747
QY		5305	ACAGTGCACCTGTAGCTATCTAFTTACTACAAAATTTGTTCCTTCCTCCACAGAAT	5364
Db		4748	TCAGTGCACCTGTAGCTATCTAFTTACCACAAAATTTGTTCCTTCCTCCACAGAAT	4807
QY		5365	ATTGGGACTGTAAAGGCCCTTGCTCCAGAGSTGGTGTGCAGATCTCTGCCCTTACTAAACT	5424
Db		4808	ATTGGGATACCAATAAAGCCCTTACTACAGAGSTGGTGTGCAGATCTCTGCCCTTACTCAAGT	4867
QY		5425	GTTTGAAGCAAAAAAACACCCGTGGTCAAGTACCCTAGAAAAGAAATAGTTTGATAGAGC	5484
Db		4868	CITTTGAAGCAGAAAAGTGTGTGTGCAGTACCCTAGAAAAGAAATAGTTTGATAGAGC	4927
QY		5485	TTCTGTATGACTATAGCTGCTCCTGNAATCAACGTTCTCATTTTCAGTGCACAGGTCTG	5544
Db		4928	TTCTGTAGGACTACAGCTGCTTCTAAATCAGGCTTCTCACTTTAGGTGTCCACGGTCTG	4987
QY		5545	CAGATGATGAGCAAAAGCATCCTGTCTCTGCCTTTTCTGTGGGTATATCTATGTCTC	5604
Db		4988	CAGATGATGAGCAAAAGCATCCTGTCTCTGTCTTTCTGTGGGGCACTCTGTGTCTC	5047
QY		5605	AGAACATTTTCTGCCAGGAAATTTGTAACGGGAAGAGSTGTGGAGCTTGCAATTTTTCAG	5664
Db		5048	AGAACATCTCTTGCCAGAATAAGTGAATGGGAAGAGGTGTGAGCGTGCCTTTTTCATG	5107
QY		5665	CAC TTC ACT GTG GAG CGG AG CTG CAT TTT C TA A AA AT CA GA GA AT CC G AG ST GG TC	5724
Db		5108	CGCTTCATTTGTTGTGTGGAGTGTGCATTTTCTTAAAAATCCGAGAATCAGGGTGTCTC	5167
QY		5725	TGTTTGAAGTAAGCCAGAGGCTGTGCCATCCAGCTCCTTACTTGGATGAATATGGAG	5784
Db		5168	TGTTGGAAGAAAAGCCAGAGGCTGTGCCCTACCGACCCCTTACTTGGATGAATATGGAG	5227
QY		5785	AACACAGCCCTGCCTGAAAGAGGGCAACCCCTTCATTTATCTCTGTGACCGGTATCGGA	5844
Db		5228	AACACAGCCAGGGCTAAAGAGAGSAAACCCACTTCATTTATCTCTGGGAGCGGTATCGGA	5287
QY		5845	AGCTCATTTTGGTCTGGCAACAACACTTGCATTTATAGAAGATTTGCTAGSAGCCAAGAGA	5904
Db		5288	AGCTGCATTTGGTCTGGCAACACACTTGCATTTATAGAAGATTTGCTCGAGGCCAGGAGA	5347
QY		5905	CTAATCAGATGTTATTGGATTCAACTGGCAGTTACTGTGAGCTCCAACTCTGCCCTCAAG	5964
Db		5348	CTAATCAGATGCTATTGGATTTAAC TG GCAG TTACTCTGAGCTTCAGTTCTGCCCTCAAG	5407
QY		5965	CAAATCACAATGACGACAGTAGTAAGGCTGATTC AAA AT TAT G GA A A A C T T T C T G A G	6024
Db		5408	ACAATCATGATGCATCAATAAAGACTGATCTAAAATTTCTAGAGAACCTTCTGTGAGG	5467
QY		6025	GCTGGAAAAGTATTGGAGGCTTTTTCCTCCATGTCCAGGTTCACATTACATCAATAAAT	6084
Db		5468	ACGGGGGAAGTATTGGAGGCTTTTTCATCCATGTCCAGATTCCACACATTAATAAAT	5527

Qy	6085	ATTTCTTAATGAGCATATTGCTTCAATTAGCAACACATATGCTTCACAGGAAAAA-AGGAC	6143
Db	5528	ATTCCTTAATGGAATATTGCTTTCAAATTATCAACACATAAGCTTCAAGGGAAGAAACAAGAC	5587
Qy	6144	ATAGATCAAACTGT-----TTT	6160
Db	5588	ATAGATAATGTTTTATGTTCTAGAACACATAAGAAATGCTTGTTCATCCAAGTGTCAT	5647
Qy	6161	ATGTGCTAGTATTTCCAGGAATTTATTCCTCTCATAAATTTGTCCTCATTTTCATTTATTT	6220
Db	5648	TTCTGCATAATATTTCCAGAAACCTCTTCCCTTCATAACTGTCCTAGTTCATTTTCATAT	5707
Qy	6221	CATCCACTTGGTAGATCAAGTCACAGTCACAAACAGTTGTAGACATTTTATGTTGGTTAAAC	6280
Db	5708	CACCCACCTGGTTAATGAGGTACACATTAAGCATTTTGGGACATTTCTCCATCGGCTAAC	5767
Qy	6281	TCATTCTGAAATTTGTATTTGGTGTTT	6307
Db	5768	ATCTCTGCACCTTTGTATTTGGTGTTT	5794

```

RESULT 3
US-08-982-956-2
; Sequence 2, Application US/08982956
; Patent No. 5861312
; GENERAL INFORMATION:
; APPLICANT: Varshavsky, Alexander
; APPLICANT: Kwon, Yong Tae
; TITLE OF INVENTION: NUCLEIC ACID ENCODING MAMMALIAN UBRI
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kevin M. Farrell, P.C.
; STREET: P.O. Box 999
; CITY: York Harbor
; STATE: ME
; COUNTRY: US
; ZIP: 03911
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/08/982.956
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Farrell, Kevin M.
; REGISTRATION NUMBER: 35,505
; REFERENCE/DOCKET NUMBER: CIT-2001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (207) 363-0558
; TELEFAX: (207) 363-0528
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1001 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..999
; US-08-982-956-2

```

	Query Match	15.88;	Score 996.2;	DB 2;	Length 1001;
	Best Local Similarity	99.7%;	Pred. No. 5.3e-279;		
	Matches 998;	Conservative	0;	Mismatches 3;	Indels 0; Gaps 0;
Qy	2799	ATGATCCCAATAAGTCTTTGTACTGGTACTTCAGAGGTATGAAC	TTGCCGAGGCTTTT	2858	
Db	1	ATGATCCCAACAAGTCTTTGTACTGGTACTTCAGAGGTATGAAC	TTGCCGAGGCTTTT	60	



Db 481 GCATTGCGCCACCACTCTGAAATCTGCCCTGTTTCAGCAAGTATTAACTT 540  
QY 3339 CTCACACTGTGATATCATATGATGATCTTCAGACCGTATTGAGCGGCAATAGACACA 3398  
Db 541 CTCACACTGTGATATCATATGATGATCTTCAGACCGTATTGAGCGGCAATAGACACA 600  
QY 3399 GATTCCTAACTGTGTGACCGAAGGATGCTTCCAAATGGCTTTTCATATCTGCGATTGGGT 3458  
Db 601 GATTCCTAACTGTGTGACCGAAGGATGCTTCCAAATGGCTTTTCATATCTGCGATTGGGT 660  
QY 3459 TTACTAGAGAGAACACACAGCTTCAAAAGCTCTGAGAGAGAAAGTAACATTTGACTTT 3518  
Db 661 TTACTAGAGAGAACACACAGCTTCAAAAGCTCTGAGAGAGAAAGTAACATTTGACTTT 720  
QY 3519 TATCATAGGCTTCAAGATTGGGAAGTTTCAGCCATGAATATACAAATGCTTTTGGAAAAA 3578  
Db 721 TATCATAGGCTTCAAGATTGGGAAGTTTCAGCCATGAATATACAAATGCTTTTGGAAAAA 780  
QY 3579 CTCAAAGGAATTCCTCAGTTAGAGCCAGAGGACATGATACGTTGGATCTTCAGATG 3638  
Db 781 CTCAAAGGAATTCCTCAGTTAGAGCCAGAGGACATGATACGTTGGATCTTCAGATG 840  
QY 3639 TTTGACACAGTGAACGGATTAGAGAAAAATCTTGTAAATTTAGTACCAACCATCAGGA 3698  
Db 841 TTTGACACAGTGAACGGATTAGAGAAAAATCTTGTAAATTTAGTACCAACCATCAGGA 900  
QY 3699 TCGGAATCTATTAAAGATGATGAGATTAATCTATGATAAGAAAAAGCAAGCAAAAAA 3758  
Db 901 TCGGAATCTATTAAAGATGATGAGATTAATCTATGATAAGAAAAAGCAAGCAAAAAA 960  
QY 3759 AAAGCTGAAGCTGTAGGCTATCATGCCAGAGAGATCATGGC 3799  
Db 961 AAAGCTGAAGCTGTAGGCTATCATGCCAGAGAGATCATGGC 1001

RESULT 5

US-08-232-463-14/c  
; Sequence 14, Application US/08232463  
; Patent No. 5670367  
; GENERAL INFORMATION:  
; APPLICANT: DORNER, F.  
; APPLICANT: SCHEIFLINGER, F.  
; APPLICANT: FALKNER, F. G.  
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 1800 Diagonal Road, Suite 500  
; CITY: Alexandria  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22313-0299  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/232,463  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIORITY APPLICATION DATA:  
; APPLICATION NUMBER: US/07/935,313  
; FILING DATE:  
; APPLICATION NUMBER: EP 91 114 300.6  
; FILING DATE: 26-AUG-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BENT, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703)836-9300

; TELEFAX: (703)683-4109  
; TELEX: 899149  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 7218 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; CLONE: pTZgpt-Fls  
; US-08-232-463-14

Query Match 0.9%; Score 57.4; DB 1; Length 7218;  
Best Local Similarity 9.4%; Pred. No. 1.2e-05;  
Matches 43; Conservative 219; Mismatches 195; Indels 0; Gaps 0;

QY 2862 AACACATATCTACAAAGACCAGGATTTGATTAACAATAATATACACTAATAGAA 2921  
Db 1481 AATTACCTATCTATGCAAGTAGTTAAAGAGATAGAAGAAATTTGTACRRRRRRRR 1422  
QY 2922 ATGCTTCAGCTCCTCATCTATATTGTTGGTGAGGTTATGTACCTGGAGTGGAAATGTG 2981  
Db 1421 RRR 1362  
QY 2982 ACCAAAGAGAGGTCAACAATCAGAGAAATCATTCACCTTGTCTTGCATTGAACCCATGCCA 3041  
Db 1361 RRR 1302  
QY 3042 CACAGTCCCATTCGCCAAAATTTACCTCAGAGAAATGAAATGAACTGGCTTAGAGAT 3101  
Db 1301 RRR 1242  
QY 3102 GTCATAACAAAGTGGCCACATTTAAGAAACAGCGTGTATCAGGCCATGGAGTTTATGAA 3161  
Db 1241 RRR 1182  
QY 3162 CTAAAGATGAATCACTGAAAGACTTCAATATGTACTTTTATCATTTACTTCAAAACCCAG 3221  
Db 1181 RRR 1122  
QY 3222 CATAGCAAGCTGAACATATGCGAGAAAGAGAGAAACAAAGAAACAAAGATGAGCA 3281  
Db 1121 RRR 1062  
QY 3282 TTGCCGCCACACCTCTGATTTCTGCCCTGCTT 3318  
Db 1061 AAGCTCCCTCGACCTGCAGCCAGCTCGGAATTAATT 1025

RESULT 6

US-08-232-463-14  
; Sequence 14, Application US/08232463  
; Patent No. 5670367  
; GENERAL INFORMATION:  
; APPLICANT: DORNER, F.  
; APPLICANT: SCHEIFLINGER, F.  
; APPLICANT: FALKNER, F. G.  
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 1800 Diagonal Road, Suite 500  
; CITY: Alexandria  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22313-0299  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/232,463



; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; CLONE: pTZgpt-Fls
; IMMEDIATE SOURCE:
;
US-08-232-463-14

Query Match
Best Local Similarity 10.1%; Score 52.2; DB 1; Length 7218;
Matches 48; Conservative 218; Mismatches 211; Indels 0; Gaps 0;

QY 4698 CTATTGGAGATGAAGAAACCTGTTTGAGGACCTTCAAAATAGGCAGCATAATGGT 4757
DB 971 CCATCAGTGGCTACTATACTATTTTCCTCGTTGCCATACGCTCACAGAAATAA 1030
QY 4758 CTGAAGACATTAATGCAGTTGTCAGTTGCACAGAGGATACCTGCTCAGGCTCGTATA 4817
DB 1031 TTCCGAGCTGGCTGCAGGTCGAGGAGCTTGCATGATGTTGCTGCTGCTGCTGCTGCT 1090
QY 4818 CAGAAACATCTGTTGCTGCTTCTATCAGTTGTTCTTCTTCAATAAATAAGAGATACA 4877
DB 1091 YY 1150
QY 4878 CCATGCTCTGCTATAGATCTGTTTCACTGTTTGGTGGTCTGCTGCTGCTGCTGCTGCT 4937
DB 1151 YY 1210
QY 4938 TCCTTGTATGGATGACCTGTTGATCTGCAGCCTTCTCAGTGTAGTCTTCTCTATAAC 4997
DB 1211 YY 1270
QY 4998 CACCTTTATCTTCCATTGATCACCATGGCAGCATGCTTCAGATACCTACTACAGTA 5057
DB 1271 YY 1330
QY 5058 GACACAGGCTACCCCTGCTCAGTTCAGAGATCAAGAGACAGTGAAGAGGCTATTCGATCT 5117
DB 1331 YY 1390
QY 5118 TCTTTCTTGCAGAAATTTCTCAATATACAAGTGGCTCCATGTTGGTGTGATATTCCT 5174
DB 1391 YY 1447

RESULT 7
US-09-754-250-3
; Sequence 3, Application US/09754250
; Patent No. 6376225
; GENERAL INFORMATION:
; APPLICANT: WEI, Ming-Hui et al
; TITLE OF INVENTION: ISOLATED HUMAN PHOSPHODIESTERASE
; TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN
; TITLE OF INVENTION: PHOSPHODIESTERASE PROTEINS, AND USES THEREOF
; FILE REFERENCE: CL001063
; CURRENT APPLICATION NUMBER: US/09/754,250

; CURRENT FILING DATE: 2001-01-05
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 111282
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc\_feature
; LOCATION: (1)..(111282)
; OTHER INFORMATION: n = A,T,C or G
US-09-754-250-3

Query Match
Best Local Similarity 0.7%; Score 45.6; DB 4; Length 111282;
Matches 57; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 293 TGTATACACACCCCTGGAAACACCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 352
DB 59380 TGTCTCCAGCTCCTGGCACCACCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 59439
QY 353 GCTGGATCCCGAGCTT 368
DB 59440 GATACCTCAAGTAATT 59455

RESULT 8
US-09-138-024-21
; Sequence 21, Application US/09138024A
; Patent No. 6004779
; GENERAL INFORMATION:
; APPLICANT: Bradley, John D.
; APPLICANT: Thompson, Craig M.
; APPLICANT: Moore, Jeffrey B.
; APPLICANT: Wobbe, C. Richard
; APPLICANT: Healy, Judith M.
; APPLICANT: Donnelly, Caroline E.
; TITLE OF INVENTION: REGULATED GENE EXPRESSION IN YEAST
; FILE REFERENCE: 0342/1D469US1
; CURRENT APPLICATION NUMBER: US/09/138,024A
; CURRENT FILING DATE: 1998-08-21
; EARLIER APPLICATION NUMBER: 60/056,719
; EARLIER FILING DATE: 1997-08-22
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 21
; LENGTH: 7333
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Plasmid pZM197
US-09-138-024-21

Query Match
Best Local Similarity 0.7%; Score 44.4; DB 3; Length 7333;
Matches 81; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

QY 989 TTGTGGAGGGGTTTTCAAAGTGGAGAGACAACTATTCTTGCAGGGATTGTGCAATTGA 1048
DB 3222 TTGTGGAGGGAATTCAAATAGGGGAACCCCTTGTATAGGTGCTATGAGTGTGGTGGCA 3281
QY 1049 TCCAACTGTGTACTCTGTATGAGTGTCTCCAGGACAGTGTTCATAAAAAATCATCGTTA 1108
DB 3282 TGATCTGTGTCTGTTGTATTTTATTTTAAATCCAAAGATCATGTGAATCATCATCT 3341
QY 1109 CAAGATGCATCTTCTACTGGA 1130
DB 3342 TTGTACCGATATATCTACTGAA 3363

RESULT 9
US-09-404-066-21
; Sequence 21, Application US/09404066

US-09-573-322-21

US-09-146-053-4

```
; Sequence 4, Application US/09146053A
; Patent No. 6399349
; GENERAL INFORMATION:
; APPLICANT: Ryan, James W.
; APPLICANT: Sprinkle, Terry Joe Curtis
; APPLICANT: Venema, Richard C.
; TITLE OF INVENTION: Human Aminopeptidase P Gene
; FILE REFERENCE: MCG103
; CURRENT APPLICATION NUMBER: US/09/146,053A
; CURRENT FILING DATE: 1998-09-02
; EARLIER APPLICATION NUMBER: 60/057,854
; EARLIER FILING DATE: 1997-09-02
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 4
; LENGTH: 50000
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-146-053-4

Query Match      0.7%; Score 42.8; DB 4; Length 50000;
Best Local Similarity 87.0%; Pred. No. 0.7;
Matches 47; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 299 CACACCCCTGGAAACACCACCATCTACTTCTGTCTATGAATTTGACTACTCTA 352
DB 23136 CAGCCCTGGCAACACCACCATCTACTTCTGTCTATGAATTTGACTACTCTA 23189

RESULT 13
US-08-991-789A-236/c
; Sequence 236, Application US/08991789A
; Patent No. 6225054
; GENERAL INFORMATION:
; APPLICANT: Frudakis, Tony N.
; APPLICANT: Smith, John M.
; APPLICANT: Reed, Steven G.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TREATMENT AND DIAGNOSIS OF BREAST CANCER
; NUMBER OF SEQUENCES: 292
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed IP Law Group
; STREET: 701 Fifth Avenue, Suite 6300
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/991,789A
; FILING DATE: 11-Dec-1997
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Potter, Jane E. R.
; REGISTRATION NUMBER: 33,332
; REFERENCE/DOCKET NUMBER: 210121.419C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 236:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 325 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; SEQUENCE DESCRIPTION: SEQ ID NO: 236:

```

```
US-08-991-789A-236

Query Match      0.7%; Score 42.2; DB 3; Length 325;
Best Local Similarity 79.4%; Pred. No. 0.041;
Matches 50; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 298 ACACACCCCTGGAAACACCACCATCTACTTCTGTCTATGAATTTGACTACTCTAGCTGG 357
DB 283 AAAAGCCCTGGCAACACCACCATCTATTTTGTGTTCTATGAAATTGACTACTCTAGGTAC 224
QY 358 ATC 360
DB 223 CTC 221

RESULT 14
US-09-062-451-236/c
; Sequence 236, Application US/09062451
; Patent No. 6344550
; GENERAL INFORMATION:
; APPLICANT: Frudakis, Tony N.
; APPLICANT: Smith, John M.
; APPLICANT: Reed, Steven G.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TREATMENT AND DIAGNOSIS OF BREAST CANCER
; NUMBER OF SEQUENCES: 297
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/062,451
; FILING DATE: 04-APR-1997
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.419C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 236:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 325 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; SEQUENCE DESCRIPTION: SEQ ID NO: 236:
US-09-062-451-236

Query Match      0.7%; Score 42.2; DB 4; Length 325;
Best Local Similarity 79.4%; Pred. No. 0.041;
Matches 50; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 298 ACACACCCCTGGAAACACCACCATCTACTTCTGTCTATGAATTTGACTACTCTAGCTGG 357
DB 283 AAAAGCCCTGGCAACACCACCATCTATTTTGTGTTCTATGAAATTGACTACTCTAGGTAC 224
QY 358 ATC 360
DB 223 CTC 221

```

Job time : 277.924 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 27, 2003, 09:29:56 ; Search time 986.772 Seconds  
(without alignments)  
15914.219 Million cell updates/sec

Title: US-09-724-126a-1  
Perfect score: 6308  
Sequence: 1 gccagaattcgacgagg.....aattttgtatttggtttt 6308

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1678620 seqs, 1244745471 residues

Total number of hits satisfying chosen parameters: 3357240

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications\_NA.\*

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*
- 2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq.\*
- 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*
- 5: /cgn2\_6/ptodata/2/pubpna/PCTUS\_NEW\_PUB.seq.\*
- 6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq.\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*
- 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq.\*
- 10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq.\*
- 11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq.\*
- 12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*
- 13: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq.\*
- 14: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq.\*
- 15: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*
- 16: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*
- 17: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	813.6	12.9	3327	10	US-09-529-063-56 Sequence 56, Appl
2	813.6	12.9	3327	12	US-10-414-378-56 Sequence 56, Appl
3	813.6	12.9	3502	10	US-09-529-063-57 Sequence 57, Appl
4	813.6	12.9	3502	12	US-10-414-378-57 Sequence 57, Appl
5	438.4	6.9	505	11	US-09-918-995-36721 Sequence 532, App
6	250.6	4.0	972	9	US-09-822-849A-532 Sequence 7, Appli
c 7	219.8	3.5	4573	13	US-10-071-766-7 Sequence 260533,
c 8	212.2	3.4	578	13	US-10-027-632-260533 Sequence 490, App
c 9	105.6	1.7	1461	9	US-09-925-301-490 Sequence 14952, A
c 10	89.6	1.4	326	10	US-09-960-332-14952 Sequence 14952, A
11	65	1.0	465	11	US-09-918-995-27470 Sequence 8907, Ap
12	60	1.0	60	12	US-09-908-975-8907 Sequence 8115, Ap
13	53.8	0.9	611	13	US-10-027-632-8115 Sequence 204690,
14	49.6	0.8	624	13	US-10-027-632-204690 Sequence 143140,
15	49.4	0.8	781	13	US-10-027-632-143140 Sequence 143138,
16	49.4	0.8	804	13	US-10-027-632-143138

17	49.4	0.8	804	13	US-10-027-632-143139	Sequence 143139,
18	49.2	0.8	572	13	US-10-027-632-198023	Sequence 198023,
19	47.4	0.8	754	13	US-10-027-632-32006	Sequence 32006, A
20	47.4	0.8	754	13	US-10-027-632-32007	Sequence 32007, A
21	47.4	0.8	2355	13	US-10-027-632-110629	Sequence 110629,
22	47.4	0.8	2355	13	US-10-027-632-111897	Sequence 111897,
c 23	47.2	0.7	538	13	US-10-027-632-91442	Sequence 91442, A
24	47.2	0.7	538	13	US-10-027-632-91443	Sequence 91443, A
25	47.2	0.7	538	13	US-10-027-632-91444	Sequence 91444, A
26	47.2	0.7	538	13	US-10-027-632-91445	Sequence 91445, A
27	47.2	0.7	1265	13	US-10-027-632-226452	Sequence 226452,
28	47.2	0.7	1265	13	US-10-027-632-226453	Sequence 226453,
29	45.6	0.7	3719	13	US-10-027-632-250210	Sequence 250210,
30	45.6	0.7	3719	13	US-10-027-632-250211	Sequence 250211,
31	45.6	0.7	111282	13	US-10-094-989-3	Sequence 3, Appli
32	45.4	0.7	598	13	US-10-027-632-242712	Sequence 242712,
c 33	45.2	0.7	768	13	US-10-027-632-154032	Sequence 154032,
34	45.2	0.7	812	13	US-10-027-632-172399	Sequence 172399,
35	45.2	0.7	11722	12	US-09-855-612-1	Sequence 1, Appli
36	45	0.7	2146	13	US-10-027-632-97126	Sequence 97126, A
37	45	0.7	2146	13	US-10-027-632-97127	Sequence 97127, A
38	45	0.7	143068	10	US-09-967-768A-316	Sequence 316, App
39	44.8	0.7	598	13	US-10-027-632-301730	Sequence 301730,
40	44.8	0.7	598	13	US-10-027-632-301731	Sequence 301731,
41	44.8	0.7	598	13	US-10-027-632-301732	Sequence 301732,
42	44.8	0.7	604	13	US-10-027-632-35389	Sequence 35389, A
43	44.8	0.7	604	13	US-10-027-632-35390	Sequence 35390, A
44	44.8	0.7	604	13	US-10-027-632-35391	Sequence 35391, A
45	44.8	0.7	604	13	US-10-027-632-81188	Sequence 81188, A

ALIGNMENTS

RESULT 1  
US-09-529-063-56  
; Sequence 56, Application US/09529063  
; Patent No. US20020102542A1  
; GENERAL INFORMATION:  
; APPLICANT: FUKUSHIMA, DAIKICHI  
; APPLICANT: SHIBAYAMA, SHIRO  
; APPLICANT: TADA, HIDEAKI  
; TITLE OF INVENTION: POLYPEPTIDE, CDNA ENCODING THE POLYPEPTIDE, AND USE OF  
; FILE OF INVENTION: THE BOTH  
; FILE REFERENCE: Q58769  
; CURRENT APPLICATION NUMBER: US/09/529.063  
; CURRENT FILING DATE: 2000-04-07  
; PRIOR FILING DATE: 1998-10-06  
; PRIOR APPLICATION NUMBER: PCT/JP98/04514  
; PRIOR FILING DATE: 1997-10-07  
; NUMBER OF SEQ ID NOS: 117  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 56  
; LENGTH: 3327  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-529-063-56

Query Match 12.9%; Score 813.6; DB 10; Length 3327;  
Best Local Similarity 55.4%; Pred. No. 7.9e-215;  
Matches 1856; Conservative 0; Mismatches 1409; Indels 87; Gaps 11;  
QY 2632 TGGAAATACCTTTACGTTGCTGTGTTGGTGTGCCAGGTGTTGCTGAGATGTGGGAA 2691  
DB 8 TAGAACACCTCTTAGATGCTTGTGTGTGCCAAGTACATGCGGAAATGGAGAA 67  
QY 2692 GAAATGACATGCTCTCTTATTAGCAGCATGTTTATACCAAGATGTTAAAGTCGACAGAA 2751  
DB 68 GAATGGTCTCTCTAGTAACACGATTTATTACTACCAATAGTGAATGACAGACGTG 127  
QY 2752 AAATGATGATAAGATATCATCATGCTTCAGATGTTGTCATCTTTATGATGCCAATA 2811

Db 128 AGATGTTTGACAGGATGTAGTAATGCTTCAGACAGGTGTCTCCATGATGGATCCAAATC 187  
QY 2812 AGTCTTGTTACTGGTCTCAGAGGTATGAACCTGCCGAGGCTTTTAAACAAGACCATAT 2871  
Db 188 ATTTCGTGATCATGCTCAGCGCGTTTGAACCTTTATCAGATTTTTCAGTACTCCAGACT 247  
QY 2872 CTACAAAA-----GACCAGGATTTGATTAAACAATATAATAC 2910  
Db 248 ATGAAAAAGATTAGTTCTGAGATTACCCATAGGATGTTGTTCAGCAACAATACTC 307  
QY 2911 TAATAGAAGAAATGCTTCCAGTCTCATCTATATTTGGTGGTGAGCGTTATGTACTCGGAG 2970  
Db 308 TAATAGAAGAAATGCTTATACCTCATTAATGCTTTGTTGGAGAGATTTAGTCCCTGGAG 367  
QY 2971 TGGAAATGTGACCAAGAGAGGTCACATGAGAGAAATCATCTCACTTGTCTTTGCAATG 3030  
Db 368 TTGGACAGGTAAATGCTACAGATGAATCAAGCGAGAGATTATCCATCAGTTGAGTATCA 427  
QY 3031 AACCCATGCCACAGTGGCATTTGCCAAAAATTTTACTTGAGATGAATAATGAACACTG 3090  
Db 428 AGCCTATGGCTCATAGTAATTTGGTAAAGTCTTTTACCTGAAGATGAGAAACAAGAGACTG 487  
QY 3091 GCTTAGAGATGTCATAAACAAGTGGCCACATTTAAGAAACCAAGTGTTATCAGGCCATG 3150  
Db 488 GCATGGAGAGTGTATCGAAGCAGTTGCCCATTTCAAGAAACCTGGATTACAGGACGAG 547  
QY 3151 GAGTTTATGAACATAAAGATGAATCACTGAAAGACTTCAATATGTAAGTCTTTATCATTA 3210  
Db 548 GCATGATGAACATAAAGCAGATGTGCCAAAGATTTCAACTTGTATTTCTATCACATTT 607  
QY 3211 CAAAACCCAGATAGCAGGCTGAACATATGCAAGAAAGAAAGAGAAACAACAAACA 3270  
Db 608 CAAGGCGAGAACAGTCCAAAGCGAAGAAAGCGCAAGAAATGAAAGAACAAATAGAG 667  
QY 3271 AAGATGAAGCATTTGCCGCCACCACCTCCTGTAATTTCTGCCCTGCTTCAGCAAAATGA 3330  
Db 668 AAGATCAGACCTCCACACCTCCGGTGTGGCTCCATTTCTGCCCTCTGTTTGCAGGCTGG 727  
QY 3331 TTAACCTTCTCAACTGTGTATCATGATGATACATTTCTCAGGACCGTATTTGAGCGGCA 3390  
Db 728 TTAACATTTTGCAGTCAGATGTCATGTTGTGCATCATGGCAACAATTTCTGCAATGGGCTG 787  
QY 3391 TAGACACAGATTTTAAGTGTGGACCGAAGGATGCTCCAAAGTGTTCATATTTCTG 3450  
Db 788 TGGAAACATAATGGATATGCTGTGTCAGAGTCCATGCTGCAAAAGGTTTACATTTAATG 847  
QY 3451 CATTTGGGTTTACTAGAGAGAGCAACAGCTTCAAAAAGCTCTCTGAAG---AACAAGTAA 3507  
Db 848 GCATGGCCTACAGAAGAAACAAACATTTAGAGAATGTCACGGAAGAGCATGTAGTAA 907  
QY 3508 CATTTGACTTTTATCATAGGCTTCAAGATTGGGAAGTTGAGCCATGAAT-----A 3558  
Db 908 CATTTACTTTCACAGAGATATCAAAACCTGTGAGCGCCAAAANAATCTCCTAGCA 967  
QY 3559 TACAAATGCTTTTGGAAAACTCAAGGAATTTCCCGAGTTAGAGGCGCAGAGACATGA 3618  
Db 968 TACTAGCTATGCTGGAACACTACAAAATGCTCCCTACCTAGAGAGTCCACAAAGACATGA 1027  
QY 3619 TAACGTGGATCTACATGTTTGACACAGTGAAGCGATTAAAGAGAAAAATCTTGTTAA 3678  
Db 1028 TTCCGTGGATTTGAAGACTTTTAAATGCTGTAAAAAGATGAGGGAGA-----GTTAC 1081  
QY 3679 TTGTAGCAACCATCAGGATCGGAATCTTAAAGATGATGAGATTACTCATGATAAG 3738  
Db 1082 CTACCAGTCCGTGGCAGACAGCAAGAACCATTAATGGAGAGAGTTCAAGGGCAAG 1141  
QY 3739 AAAAGCAGAACGAAAAAGAAAGCTGAAGCTGCTAGGCTACATCGCCAGAGATCATGG 3798  
Db 1142 ACAAGCTGAGAGGAAGAAAAAGCAGAGATTGCGACACTGCGCAGAGAAAAAGATCATGG 1201  
QY 3799 CTCAGATGCTGCTTACAGAAAAACTCATTTGAACATCAATAACTCATGATACATA 3858  
Db 1202 CTCAGATGCTGTAATCGAGGCACTTTTATTTGATGAAAAACAAGAACTCTTTTCAGCAGA 1261

QY 3859 CATCAGAAATGCCCTGGGAAAGAGATTCCATATTGGAGAGAGAGACCCACAGTCA 3918  
Db 1262 CATTAGAACTGGATGCCCTCAACCTCTGCTGTTCTTTGATCA-----TAGCCCTGTGGCTT 1315  
QY 3919 GTGACTACTTAGAATTTGCTTTGGTCTTAACGGGGTCCATCTGTTACTGAAAAGGAGG 3978  
Db 1316 CAGATATGACACTTTACAGCACTGGGCCCCGCAACAACCTCAGGTTCTTGAACAAAGACAT 1375  
QY 3979 TGCTGAGCTGATCCTTTGGCCAAAGAACAGGAGGTGAAAATAGAAAATATATGCATGG 4038  
Db 1376 TCGTTACATGATATTGTTCAAGAGGAGCAAGAGTTAAAGTGAAGAGCGGCAATGG 1435  
QY 4039 TATATCGGCTGTGTCAGAAATCTACTGCTTTAAACCCAGACAGGAGAAACCCATAG 4098  
Db 1436 TCTTGGCAGCATTTGTTTCAGAGATCACTGTATTATCAAAAACAAGAAATTAATTC 1495  
QY 4099 AACTCTCAGGAGAGCCCTAGACCCACTTTTCATGGATCCAGACTTGGCATATGAACTT 4158  
Db 1496 AAGATCCAGAAAAATAT---GATCCATTTATTCACCCCTGATCTGTTCTGTGGAACAC 1552  
QY 4159 ATACAGGAAGCTGTGGTCTATGATGCACGAGTGTCTGGCAGAGATTTTGAAGTGG 4218  
Db 1553 ACATAGTAGCTGTGGCAGCATTTATGCATGCCCATTTGTTGCAAGAGTATTTGATCCG 1612  
QY 4219 TACAGCTGAGCTCTCAGCAGCGCATTTCAATGTTGACCTTTT-----TGACTTGG 4266  
Db 1613 TTCAAGCTAAAAGAACAGCGAAGCAACAGAGATTACGCTTACATACAGAGCTATGATGAG 1672  
QY 4267 AAGTGGAGATATCTTTGCCCTCTTTGCAATCTCTGCAATCTCTGCAATCTGATCCCATTA 4326  
Db 1673 AAAACGGAGAAATCTTTGCCCTTTGTTGAAATGCTTGAATTAATGTTATCTCTGTCG 1732  
QY 4327 TTCCTTTGCAACCTCAAAAGATTAACAGTGAAGATGCAGATGCTCTGCTGAACCTTTGA 4386  
Db 1733 TGTCTT-----CTCCAGAAATATTTTAAACAGAGTTAAATTTTCAGACCAACCAA 1786  
QY 4387 CCCTGGCAGGTGGATCAGACTGTTCTGGCCAGAAATATCAGGTTATATATATAGACATG 4446  
Db 1787 ATCTGACTCAGTGGATTAGAACAATATCTCAACAATAAAGCATTACAGTTCTTATGGA 1846  
QY 4447 CTAAGAGAAACCCCAATTCCTATTTCTTTAATCAAGGAATGGAGATTTCTACTTGG 4506  
Db 1847 AAGAAGAAGTACTCCTAATATGCTCTACAAGAAATTCAGAAATGAGATTAATAC 1906  
QY 4507 AGTTCCATTCCTGAGTTTGGCGTTGAGCTTTCGATTAATTAATTAATCAATAGCATCA 4566  
Db 1907 AGCTCCCTGAAGGTTTCAGGCTGATTTCTGCTTAAGATCCCTTATCTGAGAGCATAA 1966  
QY 4567 AGGAAATGTTTATCTCTTTGCCCAACAATTTATAGAAATGGATTGAAAGTGGCCACTG 4626  
Db 1967 AAGAAATGCTAACGACATTTGGAACTGCTACCTACAAGTGGGACTTAAAGTTTCATCCCA 2026  
QY 4627 ATGAAAGGATCTCGAGTCCCATGCTGACCTGGAGACCTGCGCTTTCATATCCAGG 4686  
Db 2027 ATGAGAGAGCTCCTGCTGTTCCCAATGTTGTTGGGTAGCTGCGCGTACACCATCAAA 2086  
QY 4687 CAATTGAAATCTATTGGGAGATGAAGGAAAACCTCTGTTTGGAGCACTTCAAAATAGGC 4746  
Db 2087 GCATAGAAGAAATTTGAGTGATGAAGATAAACCATTTGTTGGTCTTTACCTTGCAGAC 2146  
QY 4747 AGCATAATGGTCTGAAAGCATTTAATGCAGTTTGCAGTTTGCACAGAGATTAACCTGCTC 4806  
Db 2147 TGGATGACTGCTTAGGTCAATTGACGAGATTTGGCCGACACACTGGACATGGCATCAG 2206  
QY 4807 AGGTCTGATACAGAAACATCTGGTTCGTTCTATCAGTTGTTCTTCTTCAATATAAT 4866  
Db 2207 TTTCAAGTGGTGAAGGACATTTTGTAAACCTTTTGTGCACTGCTGCTTAATGACAGCC 2266  
QY 4867 CAGAAGATACACATGCTCTGCTATAGATCTGTTTTCATGTTTGGTGGGTGCTGTGT 4926  
Db 2267 ATGAGGAACCTCCATGCATATTAGATATTGACATGTTTTCATTTATTTGGTGGGCTTGTG 2326

4927	QY	TAGATTCCCATCCCTTGTATTTGGGATGACCCCTGTGTGATCTGCAGCCCTTCTTCAGTTAGTT	4988
2327	Db	TT	2371
		TTGCATTCTCGTGGTTCAGTG-----TCAGGATTTTTTTCAGGGATCAGCC	
4987	QY	CTTCCCTATAACCACCTTTTATCTCTTCATCTTGATCACCATGCGACACATGCTTTCAGATAC	5046
	Db	TT	2431
2372	Db	TTGGCACTGGAGACCTTCACATTTTCCATCTGGTTACTATGGCACACATCATACAGATCT	5103
5047	QY	TACTTTACAGTAGAC-----AGGCCTACCCCTTCTCAGGTTCAAGAAGACAGTGAAGAGG	5163
	Db	TT	2491
2432	Db	TACTTTACTTCATGTACAGAAGAGAATGGCATGGATCAAGAAAAATCCCCCTGTGTGAAGAAG	
5104	QY	CTCATTTCCGATCTTCTTCTTTTTCGAGAAATTTCTCAATATACAAAGTGGCTCCTATGGGT	5163
	Db	TT	2551
2492	Db	AATCAGCAGTTCTGTGTTGTATATAAACACACTTACCAGTATAGCGGAAGTGCCTTGAAG	5223
5164	QY	GTGATATTCCTGGCTGGTATTTCTGGGTCTCACTGAAGATGCGATCACCCCTTATCTTTC	5283
	Db	TT	2611
2552	Db	AAATACCATCCGCTGGCATCTGTGGAGGAGTGTACAGCTGGAATCATGCCCTTCTCTGA	5283
5224	QY	GCTGTGTCGATTTGTTTTCACATATTTACTTGGGGTAACTCCGCCCTGAGGAACATGCATA	2671
	Db	TT	5343
2612	Db	AGTGTCTGCTTTATTTTTCATTACTTAAATGGAGTTCCTTCCCCACCGACATCAAG	2728
5284	QY	CCAATTTCTCAGAAGGAGTACAGTGCATCTGTAGCTATCTATCTTTACCTTACCAATTT	5403
	Db	TT	2788
2672	Db	T-----TCTTGGAAACAGCCATTTTGAACATTTATGTAGCTATCTTCCCTTCCAAACAAC	5463
5344	QY	TGTTCTGCTCTTCCAGGAATATTTGGGATACTCTAAGGCCCTTGCTCCAGAGTGGTGTG	2848
	Db	TT	5523
2729	Db	TCATTTGGCTTTTTCAGRAAATAGTGAGATATGAATTCACGTGATTTGAAGTTGGTGCC	2908
5404	QY	CAGATCCTGCTTACTTAACATGTTTGAAGCAAAAAACACCGTGGTGAGTACCCTAGAA	5583
	Db	TT	2968
2789	Db	GTAAACAGTGAAGTTAAAGATATCTAGAAGGTGAAAGAGATGCTATAAGATATCCAAAG	5643
5464	QY	AAAGAAATAGTTTGTATAGAGTTCCTGTATGACATATAGCTGCCCTCCTGAATCAAGCTCTC	3028
	Db	TT	5703
2849	Db	AATCTAACAAATTAATAACCTTCCAGAGATTACAGACGCCCTCATTAATCAAGCATCCA	3088
5524	QY	ATTTCAAGTGCCACCGTCTGCAGATGATGAGCGAAGCATCCTGTCTCTGTGCTTTCTTCT	5763
	Db	TT	3148
2909	Db	ATTTCTCGTGCCGAAATCAGGTGGTGATAAGAGCAGAGCCCCAACCTCTGTGCTTTGTGT	5823
5584	QY	GTGGGCTATATGTTCTCAGAAATTTTGTGTCAGGAATTTGTGAACGGGGAAGAGG	3208
	Db	TT	5883
2969	Db	GCGGATCTCTGCTGTCTCCAGAGTTACTGTGTGCCAGACTGAACCTGGAAGGGGAGGATG	3268
5644	QY	TTGGAGCTGTATTTTTCAGCACTTCACTGTGGCGCGAGTCTGCATTTTCTCTTCTTCTTCT	5935
	Db	TT	3320
3029	Db	TAGAGCCTGCACAGCTCACACCTACTCTGTGGCTCTGGAGTGGGCATCTTCTCTGAGAG	
5704	QY	TCAGAAATGCCAGTGGTCTCTGGTTGAAGGTAAAGCCAGAGGCTGTGCTATTCACGCTC	
	Db	TT	
3089	Db	TACGGGAATGTCAGGTGCTATTTTATAGTGGCAAAACCAAAAGGCTGTTTTTATCTCTCTC	
5764	QY	CTTACTTGGATGAATATGAGAAACAGACCTGGCTGAAGAGGGGCAACCCCTTCATT	
	Db	TT	
3149	Db	CTTACCTTGATGACTATGGGAGACCGACAGGGACTCAGACGGGGAATTCCTTTTACATT	
5824	QY	TATCTCGTGAAGCGGTATCGGAAGCTCCATTTGGTCTGCGCAACAACTGCATTTATAGAAG	
	Db	TT	
3209	Db	TATGCAAGAGCGATTCAGAAGATTCAGNAGCTCTGGCACCAACACAGATGTCTACAGAGG	
5884	QY	AGATTGCTAGGACCCAAAGACTTAATCAGATGTTATTTGGATTCAACTGGCA	
	Db	TT	
3269	Db	AAATGGCATGCACAGGAAGCAATCAGACACTGGTTCGCATTTGACTGGCA	

## RESULT 2

RESULT 2  
US-10-414-378-56

US-10-414-378-38  
; Sequence 56, Application US/10414378

; Publication No. US20030165981A1

TRANSMISSION NO. 00260070

	: GENERAL INFORMATION:
:	APPLICANT: FUKUSHIMA, DAIKICHI
:	APPLICANT: SHIBAYAMA, SHIRO
:	APPLICANT: TADA, HIDEAKI
:	TITLE OF INVENTION: POLYPEPTIDE, CDNA ENCODING THE POLYPEPTIDE, AND USE OF
:	TITLE OF INVENTION: THE BOTH
:	FILE REFERENCE: Q58769
:	CURRENT APPLICATION NUMBER: US/10/414,378
:	PRIOR FILING DATE: 2003-04-16
:	PRIOR APPLICATION NUMBER: US/09/529,063
:	PRIOR FILING DATE: 2000-04-07
:	PRIOR APPLICATION NUMBER: PCT/Jp98/04514
:	PRIOR FILING DATE: 1998-10-06
:	PRIOR APPLICATION NUMBER: JP 9-274674
:	PRIOR FILING DATE: 1997-10-07
:	NUMBER OF SEQ ID NOS: 117
:	SOFTWARE: PatentIn Ver. 2.1
:	SEQ ID NO 56
:	LENGTH: 3327
:	TYPE: DNA
:	ORGANISM: Homo sapiens
:	US-10-414-378-S6

  

Query Match      12.9% ; Score 813.6; DB 12; Length 3327;  
Best Local Similarity    55.4%; Pred. No. 7.9e-215;

	Matches 1856;	Conservative    0;	Mismatches 1409;	Indels         87;	Gaps            11;
Qy	2632	TGGAATATTCCTTTACAGTGTGGTGCTGGTGTGGCCACAGGTGGTTGGTGTGAGATCGTGC	CAAA	2691	
Dd	8	TAGAACCCTCCTTAGATGCTTGCTGTGTGCCAAGTACATGCCGGAATGTGGAAA	67		
Qy	2692	GAATGAGCTGCTCTATTATPAGCACGGTGTGTTTTATTACCAAGATGTTTAAGTGCAGACAAG	2751		
Dd	68	GAATGGGTCTCTCTAGTAACACAGATTTATTACTACCATAATGTGAATGCGACGCTG	127		
Qy	2752	AATGTATGATAAAGATATCATCATGCTTCAGATTGGTGCATCTTTAATGGATCCCCAATA	2811		
Dd	128	AGATGTTTGACAAGGATGTAGTAATGCTTCAGACAGGTGTCTCCATGATGGATGCCAAATC	187		
Qy	2812	AGTTCTGTGTACTGGTACTTCACAGGGTATGAACCTTGC CGAGGCTTTTAAACAAGCCATAT	2871		
Dd	188	ATTCTGTGATGATCATGCTCAGCGCTTGGAACCTTTATCAGATTTTCAGTACTCCAGACT	247		
Qy	2872	CTACAAAA-----GACCAGGATTTGATTAACAATAATAATATACAC	2910		
Dd	248	ATGAAAAAGATTAGTTCTGTGAGATTACCCATAAGGATGTTGTTACGACAACAATACTC	307		
Qy	2911	TAATAGAAGAAATGCTTCAGGTCTCATCTATATTGTTGGGTGAGCGTATGTACCTGGAG	2970		
Dd	308	TAATAGAAGAAATGCTATACCTCATTTATATGCTTGTGGAGAGAGATTTAGTCTCTGGAG	367		
Qy	2971	TGGGAATGFGAACAAAGAGAGGTGACATGATGAGAGAAATCATCTCACTTCTTTGCATTG	3030		
Dd	368	TTGGACAGGTAAATGCTACAGATGAATCAAAGCGAGAGATTTCATCATGTTGAGTATCA	427		
Qy	3031	AACCCATGCCACACAGTGCCTATTCGCCAAAAATTTTACCTGAGAAATGAAAAATATGAACCTG	3090		
Dd	428	AGCCTATGGCTCATAGTGAATTGGTAAGTCTTTACTGNAGATGAGACAAGGAGACTG	487		
Qy	3091	GCTTTAGAGATGCTATAAACAAAGTGGCCACATFTTAAGAAACCAGGTGTATCAGGCCATG	3150		
Dd	488	GCATGGAGAGTGAATCGAAGCAGTGTGCCCATTTCAAGAAACCTGGATTTAACAGGACGAG	547		
Qy	3151	GAGTTTATGNACTAAAGATGATCATCTGAAGACCTTCAATATGTACTTTTATCATTA	3210		
Dd	548	GCATGTATGNACTGAACACAGATGTCGCAAAAGATTCACCTTGTTATTTCTATCACTTT	607		
Qy	3211	CCAFAACCAGCATAGCAAGGCTGAACATATGCAGAGAAAGAGAGAAAACAAGAAACA	3270		
Dd	608	CAAGGCAGNACAGTCCAGGCGAGAGAACGCCNACGGAATTTGAANAAGACAAAATAGAG	667		
Qy	3271	AAGATGAAGCATTTGCCGCCACCACCACCCTCTCTGAATTTCTGCCCTGCTTTTCAGCAAAAGTGA	3330		



Db 568 AAGATACAGACTCCACACCTCCGGTGTGGCTCCATTCCTGCTTGTGGCAAGCCTGG 727  
QY 3331 TTAACCTTCTCAACTGTGATATCATGATGATACATCTCTCAGGACCGTATTTGAGCGGGCAA 3390  
Db 728 TTAACATTTTGGAGTCAGATGTCATGTTGTGTCATGATGGAACAATCTCGAATGGGCTG 787  
QY 3391 TAGACAGATTTCACTTGTGGACCGGAAGGATGCTCCAAATGGCTTTTCATATTTCTGG 3450  
Db 788 TGAACATATGATGATGCTGCTGAGAGTCCATGCTGCAAGAGGGTGTACATTTTAAATG 847  
QY 3451 CATTGGGTTTACTAGAGAGAACCAACAGCTTCAAAAAGCTCCTGAAG---RAGAAGTAA 3507  
Db 848 GCATGGCATACAGAAGAAAACAATTTAGAGAAATGTCACGGAAGAGCATGTAGTAA 907  
QY 3508 CATTTGACTTTTATCATAGGCTTCAAGTTGGGAAGTTTCAGCCATGAAT-----A 3558  
Db 908 CATTTACCTTCACAGATATCAAAACCTGGTGAAGGCCCAAAAATTTCTCCTAGCA 967  
QY 3559 TACAATGCTTTTGGAAAACTCAAGGAATTTCCCGAGTTAGAGGCCGAAGACATGA 3618  
Db 968 TACTAGCTATGCTGGAAACACTACAAAATGCTCCCTACCTAGAGTCCCAAAAGACATGA 1027  
QY 3619 TAACGTGGATCTCAGATGTTTGACACAGTGAAGCGATTAAAGAGAAAATCTGTGTTAA 3678  
Db 1028 TTCCGTGGATATTGAAGACTTTTAATGCTGTTAAAAGATGAGGGAGA-----GTTTCA 1081  
QY 3679 TTGTAGCAACACATCAGGATCGGAATCTATTAAAGATGATGAGATTACTCATGATAAG 3738  
Db 1082 CTACCACTCCGTTGGCAGACAGAGGAACCATTAATGGAAGAGTTCAAGGGACAAG 1141  
QY 3739 AAAAGCAGAACGAAAAGAAAGCTGAAGCTGCTAGGCTTACATCGCCGAAGATCATGG 3798  
Db 1142 ACAAGCTGAGAGGAAGAAAAGCAGAGATTGCCAGACTGCCGAGAGAAAAGATCATGG 1201  
QY 3799 CTGAGTGTCTGCTTACAGAAAACCTTCATTGAACTCATAACTCATGTATGACAAATA 3858  
Db 1202 CTCAGATGTCTGAAATCGCGCATTTTATTGATGAAAACAAGAACTCTTTTCAGCAGA 1261  
QY 3859 CATCAGAAATGCTTGGGAAAAGAGATTCOATTTATGAGGAGAGAGACACCCACAGTCA 3918  
Db 1262 CATTAGAACTTGGATGCGCTCAACCTCTGCTGTTTGTATCA-----TAGCCCTGTGGCTT 1315  
QY 3919 GTGACTACTCTAGAAATTTGCTTTGGGTCTTAAACGGGTCTCATCTGTTACTGAAAGGAGG 3978  
Db 1316 CAGATATGACATTTACAGCACTGGGCCCGCACAACTCAGGTTCTTGAACAAGACAAT 1375  
QY 3979 TGTGACGTGATCCTTTGGCAAGAACAGAGAGGTGAATAAGAAAATATGCCATGG 4038  
Db 1376 TCGTTACATGTATATTGTCTCAAGAGGAGCAAGAAGTTAAAGTGAAGCAGGGCAATGG 1435  
QY 4039 TATTATCGGCTGTGCCAGAAATCTACTGCTTTAACCCAGCAGAGGGGAAAACCCATAG 4098  
Db 1436 TCTTGGCAGCATTTGTTTCAGAGATCAACTGATGATTTATCAAAAACAGAAATTAATTC 1495  
QY 4099 AACTCTCAGGAGAACGCTTAGACCCACTTTTCATGGATCCAGACTTTGGCATATGGAACCT 4158  
Db 1496 AAGATCCAGAAAATAT--GATCCATTATTCATGSCACCTGATCTGTTGTGGAACAC 1552  
QY 4159 ATACAGGAAGCTGTGGTCATGTAATGACGAGGTGCTGGCAGAAATTTTGAAGCTG 4218  
Db 1553 ACACGTAGTGTGGGCATATTATGATGCCCATTTGTTGGCAAGGATTTTGTATTCGG 1612  
QY 4219 TACAGCTGAGCTCTCAGCAGCGCATTTGTTGACCTTTT-----TGACITGG 4266  
Db 1613 TTCAAGCTTAAGAACAGCAGGAAGCAACAGAGATTACGCTTACATACGAGCTATGATGAG 1672  
QY 4267 AAAGTGAGAAATATCTTTGGCCTCTTTGCAAAATCTCTGTGCAATACTGTATGCCCATTA 4326  
Db 1673 AAAACGGAGAAATCTCTTTGGCCCCCTTTGTGAATGCTTGAGTAATACGTATTTCTCTGC 1732  
QY 4327 TTCCTTTGCAACCTCAAAAGATAAACAGTGAAGATGCAAGTGCCTTTGCTCAACTTTTGA 4386  
Db 1733 TGCCTT-----CTCCAAGAAATATTTTAAACAACAGGTTAAATTTTTCAGACCAACCAA 1786

QY 4387 CCCTGGCACGGTGGATACAGACTGTTCTGGCCAGAAATATCAGGTTTAAATATAGACATG 4446  
Db 1787 ATCTGACTCAGTGGATTAGAACAATATCTCAGCAAAATAAAGCATTTACAGTTCTTTAGGA 1846  
QY 4447 CTAAGAGAAAACCCCAATTCCTATTCTTTTAATCAAGGAATGGAGATTTCTACTTTGG 4506  
Db 1847 AAGAAGAAAGTACTCCTTAATTAATGCCCTCTACAGAAATTCAGAAAATTTGGATGAATAC 1906  
QY 4507 AGTTCCATTCATCCTGAGTTTGGCGTTTGGCTTTCGATTTAAATATTTCAATAGCATCA 4566  
Db 1907 AGTCCCTCAGGGTTTTCAGGCTGATTTTCGCTCTAAGATCCCTTATTTCTGAGACATAA 1966  
QY 4567 AGAAATGGTTTATTTCTTTGGCCACAACAATTTATAGAAATTTGGATTTGAAGTGCACCTG 4626  
Db 1967 AAGAAATGCTAACGACATTTTGAACCTGCTACCTACAAGTGGGACATAAAGTTTATCCCA 2026  
QY 4627 ATGAAAGGATCCTCGAGTCCCATGCTGACCTGGAGCAGCTGGCTTTTCACTATCCAGG 4686  
Db 2027 ATGAAAGGATCCTCGTGTTCCTCATTAATGTTGGGGTAGCTGGCGTACACCATCCAAA 2086  
QY 4687 CAATTGAAAATCTATTGGGAGATGAAGAAAACCTCTCTTTGGAGCACTTCAAAATAGGC 4746  
Db 2087 GCATAGAAAGATTTTGGTGTGATGAAGATAAACCATTTGTTGGTCTTTTACCTTGCAGAC 2146  
QY 4747 AGCATAATGGTCTGAAAGCATTAATGCAAGTTTGGAGTTGCGACAGAGATTTACCTGCTC 4806  
Db 2147 TGGATGACTGCTTATAGGTCATTTGAGAGATTTTGGCGCAGCACACTGGACAGTGGCATCAG 2206  
QY 4807 AGGTCCTGATACAGAAACATCTGGTTCGCTCTCTATCATAGTTGTTCTTCTTCACTAAAT 4866  
Db 2207 TTTTCAGTGGTCAAGGACATTTTGTAAACCTTTTTCATCACTGCTGCTCAATGACAGCC 2266  
QY 4867 CAGAAGTACACCATGCTCTGCTATAGATCTGTTTCATGTTTGGTGGGTGGTGTGT 4926  
Db 2267 ATGAGAACTTCCATGCAATATAGATTTGACATGTTTTCATTTATTTGGTGGCTTGGTGC 2326  
QY 4927 TAGCATTTCCCATCTCTTATTTGGGATGACCTGTTGATCTGCGACCTTTCTTCACTAGTT 4986  
Db 2327 TTGCATTTCCGTTGCGTGCAGTG-----TCAGGATTTTTCAGGGATCAGCC 2371  
QY 4987 CTTCCATTAACACCTTTATCTCTTCCATTTGATCACCATGGCACACATGTTTCAGATAC 5046  
Db 2372 TTGGCACTGGAGACCTTTCACATTTTCCATCTGGTTACTATGGCACACATCATACAGAT 2431  
QY 5047 TACTTACAGTAGACAC----AGSCCTTACCCTTGTCTCAGTTCAAGAAGACAGTGAAGAG 5103  
Db 2432 TACTTACCTCATGTACAGAGAGAAATGGCATGGATCAAGAAATCCCTTGTGGAAGAG 2491  
QY 5104 CTCATTTCCGCATCTTCTTTTTCAGAAATTTCTCAATATACAAAGTGGCTCCATTTGGGT 5163  
Db 2492 AATCAGCAGTTCTTCTTTTGTATAAACAACATTTACCAGTATACGGGAAGTGCCTTGAAG 2551  
QY 5164 GTGATATTTCCGCGCTGATTTTGGGGTCTCACTGAAGAAATGGCATCAGCCCTTATCTTC 5223  
Db 2552 AAATACCATCCGGCTGGCATCTGTGGAGGAGTGTGAGAGCTGGAATCATGCTCTTCCCTGA 2611  
QY 5224 GCTGTGCTGATGTTTTCACATTTTACTTTGGGTAACCTCCGCTGAGGAACATGCATA 5283  
Db 2612 AGTGTGCTGCTTATTTTTCATTTAAATGGAGTTTCTTCCCTCCACCCGACATTCAG 2671  
QY 5284 CCAATTTCCAGAGAGGAGATACAGTGCATCTGTAGCTATCTATCTTTTACCTACAATTT 5343  
Db 2672 T---TCCTTGGACAAGCCATTTTGAACATTTTATGATGCTATCTTCCCTTACCAACAACC 2728  
QY 5344 TGTTCCTCTCTTCCAGGAATATTTGGGATCTGTAAGGCCCTTGTCTCCAGAGGTGGTGTG 5403  
Db 2729 TCATTTGCTTTTTCAGAAAAATAGTGAGATAATGAATTCACATGATTGAAGTTGGTGCC 2788  
QY 5404 CAGATCCTGCTTACTAACTGTTTGAAGCAAAAAACACCGTGGTTCAGCTACCTTAGAA 5463  
Db 2789 GTAACAGTGAAGTTAAAAAGATATCTAGAAGGTGAAGAGATGCTATAAGATATCCAAGAG 2848





## RESULT 4

US-10-414-378-57  
; Sequence 57, Application US/10414378  
; Publication No. US200301065981A1  
; GENERAL INFORMATION:  
; APPLICANT: FUKUSHIMA, DAIKICHI  
; APPLICANT: SHIBAYAMA, SHIRO  
; APPLICANT: TADA, HIDEAKI  
; TITLE OF INVENTION: POLYPEPTIDE, CDNA ENCODING THE POLYPEPTIDE, AND USE OF  
; FILE REFERENCE: Q58769  
; CURRENT APPLICATION NUMBER: US/10/414,378  
; CURRENT FILING DATE: 2003-04-16  
; PRIOR APPLICATION NUMBER: US/09/529,063  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: PCT/JP98/04514  
; PRIOR FILING DATE: 1998-10-06  
; PRIOR APPLICATION NUMBER: JP 9-274674  
; PRIOR FILING DATE: 1997-10-07  
; NUMBER OF SEQ ID NOS: 117  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 57  
; LENGTH: 3502  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: sig\_peptide  
; LOCATION: (56)..(109)  
; FEATURE:  
; NAME/KEY: mat\_peptide  
; LOCATION: (110)..(3382)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (56)..(3382)  
; US-10-414-378-57

Query Match 12.9%; Score 813.6; DB 12; Length 3502;  
Best Local Similarity 55.4%; Pred. No. 8.2e-215;  
Matches 1856; Conservative 0; Mismatches 1409; Indels 87; Gaps 11;  
QY 2632 TGAATATACCTTTACGTTCTGCTGGTGTGGTGGCCAGGTGTTGCTGAGATGTGGCGAA 2691  
DB 63 TAGAACACCTCTTAGATGCTGTCTGTGTGCCCAAGTACATGCCGGAATGGGAGAA 122  
QY 2692 GAAATGGAGTCTCTTATAGCCAGGTGTTTATACCAAGATGTTAAGTGCAGAGAG 2751  
DB 123 GAAATGGGTCTCTCTAGTAAACCAGATTTATACCAATATGTGAAATGCAGACGTG 182  
QY 2752 AAATGTATGATAAAGATATCATCTGCTTCAGATTGGTGTGATCTTTAATGGATCCCAATA 2811  
DB 183 AGATGTTTGACAAGATGTAGTAAUGCTTCAGACAGGTGCTCCATGATGGATCCAAATC 242  
QY 2812 AGTTCTTGTTACTGGTACTCAGAGGTATGAACCTGCCGAGGCTTTTAAACAAGACCATAT 2871  
DB 243 ATTTCCTGATGATCATGCTCAGCGGCTTTGAACCTTTATCAGATTTTCAGTACTCCAGACT 302  
QY 2872 CTACAAAA-----GACCAGGATTTGATTAACAATATAATACAC 2910  
DB 303 ATGGAAAGAAAGATTTAGTCTCGAGATTACCCATAAGGATGTGTGCAGCAGAAATACTC 362  
QY 2911 TAATGAAGAAATGCTTCAGGTCCCTCATCTATATTTGGTGGTGGAGCTTATGTACCTGGAG 2970  
DB 363 TAATGAAGAAATGCTATACCTCATATTAATGCTTGTGGAGAGAGATTAGTCCCTGGAG 422  
QY 2971 TGGGAAATGTGACCAAGAGAGGTGCACAAATGAGAGAAATCATCTTCTGCTTTGGCATG 3030  
DB 423 TTGGACAGATAAATGCTACAGATGAATCAAGCAGAGATTAATCCATCAGTTGAGTATCA 482  
QY 3031 AACCCATGCCACAGTGCCTATGCCAAAAATTTACCTGAGATGAATAATATGAACACTG 3090  
DB 483 AGCCTATGCTCATAGTGAATGGTAAAGTCTTTACCTGAAGATGAGAACAGGAGACTG 542

QY 3091 GCTTAGAGAAATGTCAATAAACAAGGTGCCACATTTTAAGAAACACAGGTGTATCAGGCCATG 3150  
DB 543 GCATGGAGAGTGAATCGAAGCAGTGGCCCATTTCAAGAAACCTGGATTAAACAGGACGAG 602  
QY 3151 GAGTTTATGAACATAAAGATGAATCAGTCAAGAGACTTCATATGATGACTTTTATCATTA 3210  
DB 603 GCATGTATGAACATAAAGTGAAGAGTTCAGTCAAGAGTTCAGTCAAGTTCATCACTTTT 662  
QY 3211 CCAAAACCCAGCATAGCAAGCTGAACATATCAGAGAAGAAAGAGAGAAACAGAAACA 3270  
DB 663 CAAGGGCAGAACAGTCCRAAGCAGAGAAAGCCCAACGGAAATTTGAAGACAAAATAGAG 722  
QY 3271 AAGATGAAGCATTTGCCGCCACACACCTCCTCTGAATTTCTGCCTCTGTTTTCAGCAAAAGTGA 3330  
DB 723 AAGATACAGCACTCCCACTCCGGTGTGCTCCATCTCTGCTGCTGTTTGAAGACCTG 782  
QY 3331 TTAACCTTCTCAACTGTGATATCATGATGTACATTTCTCAGGACCGTATTTTGAGCGGCGAA 3390  
DB 783 TTAACATTTTGCAGTTCAGATGTGTTGTGCATCATGGAAACAATTTCTGCAATGGCTG 842  
QY 3391 TAGACACAGATTTAACTTTGGCCGAGGAGTCTCCAAATGCTTTTCATATATCTG 3450  
DB 843 TGAACATAATGGATATGCCGTGAGAGTCCATGCTGCAAGGGTGTACATTTAATTTG 902  
QY 3451 CATTGGGTTTACTAGAAGAGAACACAGCTTCAAAAAGCTCCTGAAAG---AAGAAGTAA 3507  
DB 903 GCATGGCAGTACAGAAGAAAACAACATTTAGAGANTGTCACGGAAGAGCATGTAGTAA 962  
QY 3508 CATTGTACTTTTATCATTAAGGCTTCAAGATGGGAAAGTTTCAGCCATGAAT-----A 3558  
DB 963 CATTGTACTTCACTCAGAAAGATATCAAAACCTGGTGAAGCGCCAAAAATTTCTCTAGCA 1022  
QY 3559 TACAATGCTTTTGGAAAACTCAAGGAATTTCCCAAGTTAGAACGCCAGAGGACATGA 3618  
DB 1023 TACTAGTATGCTGGAAACACTACAAAATGCTCCTACTAGAAAGTCCACAAAGACATGA 1082  
QY 3619 TAAGCTGATACTTCAGATGTTTGACACAGTGAAGCGATTAAGAGAAAAAATCTTGTATA 3678  
DB 1083 TTCGGTGGATTTGAAGACTTTTATGCTGTTTAAAGATGAGGAGA-----GTTTCA 1136  
QY 3679 TTGTAGAACACATCAGGATCGGAATCTATTAAGAATGATGAGATTTACTCATGATAAG 3738  
DB 1137 CTACAGTCCCGTGGCAGAGACAGAAAGAACCATTAATGAAGAGAGTTTCAAGGGACAAA 1196  
QY 3739 AAAAGCAGACGAAAAAGAAAGCTGAAGCTGCTAGGCTAGCTACATGCCAGAGATCATGG 3798  
DB 1197 ACAAGCTGAGAGAAAGAAAGCAGAGATTGCCAGACTGCCAGAGAAAAAGATCATGG 1256  
QY 3799 CTCAGATGCTGCTCTTACAGAAAAAATTCATTGAAACTCATAACTCATGTATGACAAATA 3858  
DB 1257 CTCAGATGCTGAAATGCAGCGGCATTTTATGATGAACAAAGAACTTTTCAGCAGA 1316  
QY 3859 CATCAGAAATGCCCTGGGAAAGAAATCCATTATGGAGAGAGAGACCCAGCAGTCA 3918  
DB 1317 CATTAAGAACTGGATGCTCAACCTCTGCTGCTTTGATCA-----TAGCCCTGTGGCTT 1370  
QY 3919 GTGACTACTCTAGAAATGCTTTGGGTCTTAACGGGTCCATCTCTTACTGAAAGAGG 3978  
DB 1371 CAGATATGACACTTTACAGCACTGGGCCCGCCACAACTCAGGTCTCTGAAACAAGACAAT 1430  
QY 3979 TGCTGAGCTGATCCTTTGCCAAGAAACAGGAGGTGAAATAAGAAAATATATGTCATGG 4038  
DB 1431 TCCTTACATGATATTGTTCAAGAGGACGAAGATTTAAGTGAAGACGAGGCAATGG 1490  
QY 4039 TATTATCGGCTGTGTCCAGAAATCTACTGCTTTAACCCAGCAGCAGGGGAAACCCATAG 4098  
DB 1491 TCTTGGCAGCATTTGTTTCAGAGATCAACTGATTTATCAAAAAACAGAGTAAATTTATTC 1550  
QY 4099 AACTCTCAGGAGAGACCCCTAGACCCACTTTTCATGGATCCAGACTTGGCATATGACAT 4158  
DB 1551 AAGATCCAGAAAAATAT---GATCCATTTATCATGCACCTGATCTGTCTTGTGGAACAC 1607

QY 4159 ATACAGGAAGCTGTGTCATGTAATGCACGAGTGTCTGCCAGAAAGTATTTTCAAGCTG 4218  
Db 1608 AACTAGTAGCTGTGGGACATTAATGATGCCATCCCATTTGTCGAAGATATTTGATTCGG 1667  
QY 4219 TACAGCTGAGCTCTCAGCAGCGGATTCATGTTGACCTTTT-----TCAGTTGG 4266  
Db 1668 TTCAAGCTAAGACACGAGGACACAGAGATTACGCTTACATACAGAGCTATGATGTAG 1727  
QY 4267 AAGTGGAGAAATATCTTTGGCCCTTTGCAAAATCTGTGCAATPACTGTGATCCCCATTA 4326  
Db 1728 AAAACGGAGAAATCTTTGGCCCTTTGCAATCTGTGAAATCTGTGAAATPACTGTGATTCCTCTGC 1787  
QY 4327 TTCTTTGCAACCTCAAAAGATAAACAGTGAAGATGAGATGCGATGCTTCTGCTCAACTTTTGA 4386  
Db 1788 TGCTTT-----CTCCAAGAAATATTTTAAACACAGGTTAAATTTTTCAGACCAACCAA 1841  
QY 4387 CCCTGGCAGCGGTGATACAGACTGTTCTGCGCAAGATATACAGGTTAATAATATAAGACATG 4446  
Db 1842 ATCTGACTAGTGGATTAGAACATATCTCAGCAAAATTAAGCAATTTACAGTTCTTTAGGA 1901  
QY 4447 CTAAGGAGAAAACCCAAATTCCTATTTCTTTTAAATCAAGGAATGGGAGATTCTACTTTGG 4506  
Db 1902 AAGAAGAAAGTACTCTTAATAATGCCCTTACAAAGAAATTCAGAAATGTGGATGAATTAC 1961  
QY 4507 AGTTCCATTCCACTCCTGAGTTTGGCGTTGAGTCTTCGATTAATATTAATAGCATCA 4566  
Db 1962 AGCTCCCTGAAGGTTTCAGGCGTGAATTTCTGCTAAGATCCCTTATTTCTGAGAGCATAA 2021  
QY 4567 AGGAATAGGTTATTTCTTTGCGCAACAATAATTAAGAATTTGGAATGAAAGTGCCACTG 4626  
Db 2022 AAGAAATGCTAACGACATTTGGAATCTCTACCTACAGGTGGGACTTAAGGTTTCATCCCA 2081  
QY 4627 ATGAAGGAGTCTCTGAGTCCCATGCTGACCTGAGCACTCTGCTTTTCTACTATCCAGG 4686  
Db 2082 ATGAAGAGGATCCTCGTTCCTAATAATGTTGGGTAGTCTGCGCTGACACCATCCAAA 2141  
QY 4687 CAATTGAAATCTATTGGGAGATGAAGAAACCTCTGTTTGGAGCAGCTTCAAAATAGGC 4746  
Db 2142 GCATAGAGAAATTTTGAGTGATGAAGATAAACCATTGTTGGTCTTTTACCTTTGCAGAC 2201  
QY 4747 AGCATAATGTTCTGAAAGCATTAATGAGTTCAGTTCAGTGCAGAGAGATTACCTGTCTC 4806  
Db 2202 TGGATGACTGTTTAGTCAATGACGAGATTGCGCGAGCACATGACAGTGGCATCAG 2261  
QY 4807 AGTTCCTGATACAGAAACATCTGGTCTCTTCTATCAGTTGTTCTTCTTCAACATAAAT 4866  
Db 2262 TTTTCAGTGGTCAAGGACATTTTGTAACTTTTTCATCAGTCTGCTGCTTAATGACAGCC 2321  
QY 4867 CAGAAATACACCATGCTCTGCTATAGATCTGTTTCATGTTTGGTGGGTGCTGTGT 4926  
Db 2322 ATGAGGAACCTCCATGATATAGATATTGACATTTTTCATTTATTTGGTGGGCTTGTGTC 2381  
QY 4927 TAGCATTCCTCCTCTGTTTGGATGACCTGTTGATCTCAGCCTTCTTCTCAGTTAGTT 4986  
Db 2382 TTGCATTTCTCGTTCAGTG-----TCAGGATTTTCAGGGATCAGCC 2426  
QY 4987 CTTCCTATAACCCATTTATCTCTCCATTTGATCACCATTGGACACATGCTTCCAGATAC 5046  
Db 2427 TTGGCACTGGAGACCTTCACATTTTCCATCTGTTTACTATGGCACACATCATACAGATCT 2486  
QY 5047 TACTTTACAGTAGACAC---AGGCCATCCCTTGTCTAGGTTCAAGAGACAGTGAAGAGG 5103  
Db 2487 TACTTTACCTCATGTACAGAGAAATGGCATGGATCAAGAAATATCCCTCTTGTGAAGAAG 2546  
QY 5104 CTCATTCGGCATCTTCTTTTCAGAAATTTCTCAATATACAAAGTGGCTTCCATGGGT 5163  
Db 2547 AATCAGCAGTTCTTCTTTGTTGTAATAAACACTTCACCAGATATACGGGAAGTGCCTTGAAG 2606  
QY 5164 GTGATATTCCTGGCTGTTTGGGTCTCAGTGAAGAAATGGCATACCCCTTTATCTTTC 5223  
Db 2607 AATACCATCCGGCTGGCATCTGTGGAGGAGTGTGAGAGTGAATCATGCTTCTCTCA 2666  
QY 5224 GCTGTGTCATGTTTTTTCACATATTTTACTTTGGGGTAACTCCGCCCTGAGGAAGTCGATA 5283

Db 2667 AGTGTTCGTCTTATTTTTCATTTAAATGGAGTTCTCTTCCCCACCCGACATTTCAAG 2726  
QY 5284 CCAATTTCTGCAGAGAGAGTACAGTGCAGCTCTGTAGCTATCTATCTTTTACCTACCAATTT 5343  
Db 2727 T---TCCTTGGACAGCCATTTTGAACATTTATGTAGCTATCTTTCCCTTACCAACAACC 2783  
QY 5344 TGTTCCTCTCTCTTCAGGAATATTTGGGACTGTGAAGCCCTTGTCTCCAGAGGTGGTGTG 5403  
Db 2784 TCATTTGCTTTTTCAGAAATAAGTAGATTAATGAATTTCACTGATTGAAGTTGGTGCC 2843  
QY 5404 CAGATCCTCTTACTTAAACTGTTTGAACCAAAAAACACCCGTGGTCAAGTACCTTAGAA 5463  
Db 2844 GTAACAGTGAAGTTTAAAGATATCTAGAAGTGAAGAGATGCTATAAGATATCCAAGAG 2903  
QY 5464 AAAAGAAATAGTTTGTATAGAGCTTCTGTATGACTATAGTCTGCTCTGATTAAGCTTTCTC 5523  
Db 2904 AATCTAACAAATTAATAAACCTTCCAGAGGATTACAGACCTCTATTATCAAGCATCCA 2963  
QY 5524 ATTTCAAGTGGCCCGGCTCTGCAGATGATGAGCGAAAGCATCTGTCTCTGCTTCTTCT 5583  
Db 2964 ATTTCTCTGCGCCGAAATCAGGTGTGTATAGAGCAGAGCCCAACTCTGTGCTTGTGT 3023  
QY 5584 GTGGGCTATATCTATGTTCTCAGAACATTTGCTGCCAGGAATTTGTGAACGGGAGAGG 5643  
Db 3024 GCGGATCTCTGCTGTCTCCAGAGTTACTGCTGCCAGACTGAATCGAAGGGGAGGATG 3083  
QY 5644 TTGAGCTTTGCATTTTTCACGCACATTTCACTGTGGAGCCGAGTCTGCTGATTTTCTTAAAAA 5703  
Db 3084 TAGGAGCTGCACACCTCACACCTACTCTGTGGCTGCTGAGTGGGCACTTTCTCTGAG 3143  
QY 5704 TCAGAAATGCGGAGTGTCTGTTGAAGTTAAAGCCAGAGGCTGTGCTATCAGCTC 5763  
Db 3144 TACGGAAATGTGAGTGTCTATTTTGTGCTGCAAAACCAAGGCTGTTTTTATCTCTCTC 3203  
QY 5764 CTTACTTGGATCAATATGAGAAACAGACCCCTGGCTGAGAGGGCAACCCCTTCATT 5823  
Db 3204 CTTACTTGTGACTATGGGAGACCCGACCGGACTCAGACGGGAAATCTCTTACATTT 3263  
QY 5824 TATCTCTGAGCGGTATCGGAAGCTCCATTTGCTGTGCAACAACACATGCTATATAGAAG 5883  
Db 3264 TATGCAAGAGCGATTCAAGAGATTTCAGAACTCTGGCACCACACACAGTGTACAGAG 3323  
QY 5884 AGATTGCTAGAGCCAGACAGACTAATCAGATGTTTATTTGGATTCAACTGGCA 5935  
Db 3324 AAATTTGGACATGCACAGGAAGCCAATCAGACACTGTTGGCATTTGACTGGCA 3375

## RESULT 5

; US-09-918-995-36721  
; Sequence 36721, Application US/09918995  
; Publication No. US20030073623A1  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc.  
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
; TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES  
; FILE REFERENCE: 20411-756  
; CURRENT APPLICATION NUMBER: US/09/918,995  
; CURRENT FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US/09/235,076  
; PRIOR FILING DATE: 1999-01-20  
; NUMBER OF SEQ ID NOS: 38054  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 36721  
; LENGTH: 505  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)..(505)  
; OTHER INFORMATION: n = A,T,C or G  
; US-09-918-995-36721



```

; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20020192678A1 1330204.22
; NAME/KEY: unsure
; LOCATION: 3145-3168
; OTHER INFORMATION: a, t, c, g, or other
US-10-071-766-7

Query Match
Best Local Similarity 3.5%; Score 219.8; DB 13; Length 4573;
Matches 441; Conservative 0; Mismatches 307; Indels 17; Gaps 2;

QY 5171 TCCTGCTGGTATTTGCTGGTCTCACTGAAGAATGCAATCACCCTTATCTTCGCTGTGC 5230
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4535 TCGGGCTGGCATCTCTGGAGGAGTCTCAGAGCTGGAATCATCCCTTTCCTGAAGTTC 4476
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 5231 TGCATTGTTTTCCACTATTTACTTGGGTAACTCCGCTGAGAACTGCATACCAATTC 5290
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4475 TCGTTTATTTTTCATTAATGAGTTCCTCCCGGACGATTCAGT---TCC 4419
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 5291 TGCAGAGGAGATACAGTGCATCTGTAGCTATCTTCTTACCTACAAATTTGTTCC 5350
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4418 TGAACAAGCCATTTTGAACATTTATGTAGCTATCTTCCCTACCAACAACCTCATTTG 4359
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 5351 GCTCTTCCAGGAATATTTGGGATCTAAGGCCCTTGTCCAGAGGTGGTGCAGATCC 5410
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4358 CCTTTTCAAGAAATAGTGAATAATGAATTCACATGAAAGTTGGTCCCGTAACAG 4299
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 5411 TGCCTTACTAACTGTTTGAAGCAAAACACACCGTGGTCAAGTACCTAGAAAGAAA 5470
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4298 TGAAGTTAAAGATATCTAGAGGTGAAGAGATGCTATAAGATATCCAGAGAACTAA 4239
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 5471 TAGTTTGAATAGAGCTTCTGATGACTATAGTCCCTCCGAAATCAAGCTTCTCATTT 5530
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4238 CAAATTAATAAACCCTCCAGAGGATTACAGCACCCTCATTAATCAAGCATCCAATTC 4179
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 5531 GTGCCCCAGGTGTGAGATGATGAGGAAAGATCTGCTGCTCTGCTTTCTGTGGGG 5590
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4178 GT-----GGTGATAAGAGCAGAGGCCCAACTCTGTGCTTGTGCGGATC 4133
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 5591 TATACTATGTTTCAGAAACATTTGCTGCCAGGAATTTGAAGGGGAGAGGTTGGAGC 5650
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4132 TCGCTGTGCTCCAGAGTACTGCTGCCAGACTGAACCTGGAAGGGAGGATGTAGAGC 4073
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 5651 TTGCATTTTTCAGCACTTCACTGTGGAGCGGAGTCTGCTATTTCTTCTAAATACAG 5710
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4072 CTGCACAGCTCACACCTACTCTGCTGCTGAGTGGGATCTTCTGAGAGTACGGGA 4013
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 5711 ATGCCAGTGGTCTGTTGAAAGTAAAGCCAGAGGCTGTGCTATCCAGCTCCTTACTT 5770
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4012 ATGTCAGGTGCTATTTTGTAGCTGGCAAAACCAAGGCTGTTTTTATCTCCTCTAC 3953
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 5771 GGATGAATATGGAGAACAGACCCCTCGCTGAGAGGGGCAACCCCTTCATTTACTCG 5830
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3952 TGATGACTATGGGAGACACAGGAGCTCAGACGGGAAATCCTTTACATTTATGCAA 3893
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 5831 TGAGCGGTATCGGAAGCTCCATTTGTGCGCAACAACTGCTATATAGAGATTCG 5890
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3892 AGAGCATTCAGAAAGATTCAGAGGCTGTGGCAACACAGTGTACAGAGAAATGG 3833
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 5891 TAGGAGCCAGAGACTAATCAGATGTTATTTGGATTCAACTGGCA 5935
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3832 ACATGCACAGGAAGCAATCAGACACTGGTTGGCAITGACTGGCA 3788
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```

RESULT 8
US-10-027-632-260533
; Sequence 260533, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; POLYMORPHISMS IN THE HUMAN GENOME
; FILE REFERENCE: 108827.129
```

```

; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 260533
; LENGTH: 578
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-260533

Query Match
Best Local Similarity 3.4%; Score 212.2; DB 13; Length 578;
Matches 235; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 3838 ATAACTCATCTATGACAATACATCAGAAATGCCTGGGAAAGAGATTCATTATGGAGG 3897
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 226 ATTCACCAATAATGAAGCAAGTGACCAAAACATGTCTCATTTCTGATCATTTCTT 285
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3898 AAGAGACACCCAGCAGTCACTGACTACTCTAGAAATTCCTTGGGTCTTAAACGGGTC 3957
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 286 TTGGTAGCACCAGCAGTCACTGACTACTCTAGAAATTCCTTGGGTCTTAAACGGGTC 345
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3958 CATCTGTTACTGAAAGAGGTGCTGAGTGCATCTTTGGCAAGAAAGAGGTTGA 4017
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 346 CATCTGTTACTGAAAGAGGTGCTGAGTGCATCTTTGGCAAGAAAGAGGTTGA 405
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4018 AAATAGAAATAATGCCATGTATATCGGCTGTGTCAGAAATCTACTGCTTAAACC 4077
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 406 AAATAGAAATAATGCCATGTATATCGGCTGTGTCAGAAATCTACTGCTTAAACC 465
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4078 AGCAGAGGGGAAACCCATAGAACTCTCAGGAG 4110
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 466 AGCAGAGGGGAAACCCATAGAACTCTCAGGAG 498
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 9
US-09-925-301-490/c
; Sequence 490, Application US/09925301
; Patent No. US2002052308A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA106
; CURRENT APPLICATION NUMBER: US/09/925,301
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05882
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1694
; SOFTWARE: PatentIn ver. 2.0
; SEQ ID NO 490
; LENGTH: 1461
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1432)
; OTHER INFORMATION: n equals a,t,g, or c
```



```
; NAME/KEY: misc.feature
; LOCATION: (1452)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-301-490

Query Match      1.7%; Score 105.6; DB 9; Length 1461;
Best Local Similarity 82.7%; Pred. No. 6e-18;
Matches 134; Conservative 0; Mismatches 24; Indels 4; Gaps 1;

QY 123 AAGGCTTATCATCTGCTTCCACTATCCAAAGCTGTATGGCCACCTGCGTGC 182
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 171 ACAGGCTTATCATCTATCTCCACTTATCCAAAGCGCTAGGCCACTTCTGTGCC 112
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 183 CACCTGGAACCCAGCCAGCCCACTACTGCCCTCCACTACCACTGGTTCCTCCACCTGA 242
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 111 CAGTTGAAGCACCAGCCAGCCCACTACAGCCTCCGCTACCAACCGCTTCTCCCGCCCTGT 52
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 243 TCAGCTGCTGTGCTGCCATCTATCCCGCTTCTGCCGTGTC 284
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 51 ACAGC---TGGCGCGCCATCTGTGCCGCTCTGGCTTSC 14
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 10
US-09-960-352-14952/c
; Sequence 14952, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 14952
; LENGTH: 326
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 64-LIB3057-001-Q1-K1-H8
US-09-960-352-14952

Query Match      1.4%; Score 89.6; DB 10; Length 326;
Best Local Similarity 81.2%; Pred. No. 5.5e-14;
Matches 104; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 123 AAGGCTTATCATCTGCTTCCACTTATCCAAAGCTGTATGGCCACCTGCGTGC 182
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 142 ACAGGCTTATCGTCAATCTTCCACTTGTCCAGTAAACCGCTAGCGTACTGCTGCCA 83
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 183 CACCTGGAACCCAGCCAGCCCACTACTGCCCTCCACTACCACTGGTTCCTCCACCTGA 242
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 82 CAGCGGAGCCAGCCAGCCCACTACTGCGCTCCGCCACCCAGCGCTCTCCCGCATGA 23
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 243 TCAGCTGC 250
Db | | | | |
QY 22 GAAGCTGC 15
Db | | | | |

RESULT 11
US-09-918-995-27470
; Sequence 27470, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
```

```
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 27470
; LENGTH: 465
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(465)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-27470

Query Match      1.0%; Score 65; DB 11; Length 465;
Best Local Similarity 49.6%; Pred. No. 5.1e-07;
Matches 198; Conservative 0; Mismatches 195; Indels 6; Gaps 1;

QY 4257 TTGTGATTTGGAAAGTGAGAAATATCTTTGCCCTCTTTGCAAAATCTCTGTGCAATACTGTG 4316
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 69 TATGATGTAGAAACGGAGAAATTCCTTTGCCCTCTTTGCAATGCTTGAGTAATACTGTT 128
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 4317 ATCCCATTTATTCCTTTGCAACCTCAAAAGATAAAGAGTGCAGATGCAGATGCTCTTGCT 4376
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 129 ATTCC-----TCTGCTGCTTCTCCAGAAATATTTTAAACACAGGTTAAATTTTCA 182
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 4377 CAACCTTTTGACCCCTGGCACGGTGTATACAGATGTTCTGCGCCAGAAATATCAGGTTATAAT 4436
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 183 GACCAACCAATCTGACTCAGTGGATTTAGAACAAATATCTCAGCAAAATAAAGCATTTACAG 242
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 4437 ATAAGACATGCTAAGAGGAGAAACCAATTCCTATTTCTTTAATCAAGAAATGGAGAT 4496
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 243 TTTCTTAGGAAGAAGAAGTACTCTTAATATGCTCTACAAAGAAATTCAGAAAATGTG 302
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 4497 TCTACTTTTGGAGTTCCATTTCCATTCCTGAGTCTTTTGGGTTGAGTCTTCGATTAATATTC 4556
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 303 GATGAATTTACAGTCCCTGAGGGTTTCAGGCTGATTTTCGCTTAAGATCCCTTATCT 362
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 4557 AATAGCATCAAGGAAATGTTATTCCTTTGCGCACAAATTTATAGAATTTGAGTTGAAA 4616
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 363 GAGAGCATAAAGAAATGCTAACGACATTTTGAACCTGCTACCTACAAGTGGGCAATAAG 422
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 4617 GTGCCACTGATGAAGGAGTCTCGAGTCCCTCCCTGCTG 4655
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 423 GTTCATCCCAATGAAGAGGATCTCTGTTTCCATAATG 461
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 12
US-09-908-975-8907
; Sequence 8907, Application US/09908975
; Publication No. US20030165843A1
; GENERAL INFORMATION:
; APPLICANT: SHOSHAN, Avi
; APPLICANT: WASSERMAN, Alon
; APPLICANT: MINTZ, Eli
; APPLICANT: MINTZ, Liat
; APPLICANT: FAIGLER, Simchon
; TITLE OF INVENTION: OLIGONUCLEOTIDE LIBRARY FOR DETECTING RNA TRANSCRIPTS AND SPLI
; FILE REFERENCE: 36688-0005
; CURRENT APPLICATION NUMBER: US/09/908,975
; CURRENT FILING DATE: 2001-07-20
; PRIOR APPLICATION NUMBER: US 60/287,724
; PRIOR FILING DATE: 2001-05-02
; PRIOR APPLICATION NUMBER: US 60/221,607
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 32337
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 8907
; LENGTH: 60
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-908-975-8907
```

```
Query Match          1.0%; Score 60; DB 12; Length 60;
Best Local Similarity 100.0%; Pred. No. 2.6e-06; Indels 0; Gaps 0;
Matches 60; Conservative 0; Mismatches 0;

Qy 2939 CTATTGTGGTGAGCTTATGCTACCTGGAGTGGGAATGTGACCAAGAGGTCAC 2998
|||||
Db 1 CTATTGTGGTGAGCTTATGCTACCTGGAGTGGGAATGTGACCAAGAGGTCAC 60

RESULT 13
US-10-027-632-8115
; Sequence 8115, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8115
; LENGTH: 611
; TYPE: DNA
; ORGANISM: Human
; NAME/KEY: misc_feature
; LOCATION: (1)...(611)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-8115

Query Match          0.9%; Score 53.8; DB 13; Length 611;
Best Local Similarity 96.5%; Pred. No. 0.00082; Indels 0; Gaps 0;
Matches 55; Conservative 0; Mismatches 2;

Qy 4640 TCGAGTCCCATGCTGACCTGGAGCACCTGGCGCTTTCACCTATCCAGGCAATTGAAA 4696
|||||
Db 1 TCGAGTCCCATGCTGACCTGGAGCACCTGGCGCTTTCACCTATCCAGGCAATTGGTAA 57

RESULT 14
US-10-027-632-204690
; Sequence 204690, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
```

```
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 204690
; LENGTH: 624
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-204690

Query Match          0.8%; Score 49.6; DB 13; Length 624;
Best Local Similarity 85.9%; Pred. No. 0.012; Indels 0; Gaps 0;
Matches 55; Conservative 0; Mismatches 9;

Qy 299 CACACCCCTGGAAACCACTTCTACTTTCTGTCTATGAATTTGACTACTCTAGCTGGA 358
|||||
Db 334 CAGCCCTGGCAACCACTTCTACTTTCTGTCTATGAATTTGACTACTCTAGCTACT 393
|||||
Qy 359 TCCC 362
Db 394 TCAC 397

RESULT 15
US-10-027-632-143140
; Sequence 143140, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 143140
; LENGTH: 781
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-143140

Query Match          0.8%; Score 49.4; DB 13; Length 781;
Best Local Similarity 66.4%; Pred. No. 0.017; Indels 0; Gaps 0;
Matches 71; Conservative 0; Mismatches 36;

Qy 254 CTGCTGCCATCTATCCGCTTCTGCGCTGTTCTGAGTAATGTATACACACCTGGAAACC 313
|||||
Db 640 CTGTACTCATTAAATGCTAATGCYCCATTTCTCTATCCTCTGCTCCAGGCCCTGGTAACC 699
|||||
Qy 314 ACCATTCTACTTTCTGTCTATGAATTTGACTACTCTAGCTGGATC 360
|||||
Db 700 ACCATTCTATTTCTGTCTCTATGAATTTGACTACTCTAGCTACTTCTAGTACTTTC 746
|||||

Search completed: September 27, 2003, 20:27:06
Job time : 995.772 secs
```



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 27, 2003, 00:42:14 ; Search time 7897.46 Seconds  
(without alignments)  
19412.893 Million cell updates/sec

Title: US-09-724-126A-1

Perfect score: 6308

Sequence: 1 gccagaattcgccagcaggg.....aatttgatttggtttt 6308

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- EST:\*
- 1: em\_estba:\*
  - 2: em\_esthum:\*
  - 3: em\_estin:\*
  - 4: em\_estnu:\*
  - 5: em\_estov:\*
  - 6: em\_estpl:\*
  - 7: em\_estro:\*
  - 8: em\_hic:\*
  - 9: gb\_est1:\*
  - 10: gb\_est2:\*
  - 11: gb\_hic:\*
  - 12: gb\_est3:\*
  - 13: gb\_est4:\*
  - 14: gb\_est5:\*
  - 15: em\_estfun:\*
  - 16: em\_estom:\*
  - 17: em\_gss\_hum:\*
  - 18: em\_gss\_inv:\*
  - 19: em\_gss\_pln:\*
  - 20: em\_gss\_vit:\*
  - 21: em\_gss\_fun:\*
  - 22: em\_gss\_nam:\*
  - 23: em\_gss\_mus:\*
  - 24: em\_gss\_pro:\*
  - 25: em\_gss\_rod:\*
  - 26: em\_gss\_phg:\*
  - 27: em\_gss\_vrl:\*
  - 28: gb\_gss1:\*
  - 29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2096	33.2	2561	11 AK089616	AK089616 Mus muscu
2	826.6	13.1	960	12 BM472160	BM472160 AGENCOURT
3	806.4	12.8	3641	11 BC044903	BC044903 Mus muscu
4	771.4	12.2	865	13 BQ233617	BQ233617 AGENCOURT

5	740.2	11.7	3058	11	AK078173
6	710.4	11.3	712	13	BX104087
7	687.8	10.9	3627	11	AK083320
8	674.6	10.7	682	9	AI929033
9	668.8	10.6	898	13	BUI15953
10	664.8	10.5	668	13	BUI18516
11	656.6	10.4	756	10	BGS34574
12	637.4	10.1	641	9	AI361043
13	635.2	10.1	774	13	BUI955331
14	627.2	9.9	797	12	BG862813
15	594.8	9.4	640	14	CD368666
16	585.2	9.3	695	14	BY756516
17	583.6	9.3	703	12	BM114217
18	576	9.1	729	13	BUI703363
19	573.2	9.1	782	12	BUI086469
20	560	8.9	560	14	CB159971
21	557	8.8	565	10	BF063405
22	545.4	8.6	684	14	BY764809
23	529.2	8.4	691	14	BY736138
24	521.2	8.3	528	13	BX119443
25	517.6	8.2	659	14	BY750548
26	512.2	8.1	606	10	BE589438
27	511.4	8.1	752	14	CA313244
28	509.4	8.1	633	10	BB622408
29	501.4	7.9	506	2	HSN077106
30	495.6	7.9	777	14	CB232501
31	497.4	7.9	719	13	BUI623096
32	496	7.9	496	13	BX281065
33	495.2	7.9	583	12	BI540789
34	494	7.8	495	14	CB129142
35	487.2	7.7	1105	13	BUI751567
36	479.8	7.6	624	10	BG219270
37	477.8	7.6	481	9	AI192195
38	473.6	7.5	609	29	CC200266
39	473.4	7.5	583	10	AW971391
40	471.6	7.5	478	9	AA401319
41	469.4	7.4	473	13	BUI08567
42	464.4	7.4	584	14	CB35201
43	464.4	7.4	830	14	CB588645
44	454	7.2	522	9	AW291190
45	453.8	7.2	737	13	BUI461940

## ALIGNMENTS

RESULT 1	AK089616	2561 bp	mRNA	linear	HTC 05-DEC-2002
LOCUS	Mus musculus	activated spleen cDNA	RIKEN full-length enriched library, clone:F830005C07	product:ubiquitin protein ligase E3 component n-recognin 1, full insert sequence.	
DEFINITION	AK089616	AK089616.1	GI:26354612		
ACCESSION	AK089616	HTC; CAP trapper.			
VERSION	AK089616.1	Mus musculus (house mouse)			
KEYWORDS	HTC; CAP trapper.				
SOURCE	Mus musculus				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
AUTHORS	Carninci,P. and Hayashizaki,Y.				
TITLE	High-efficiency full-length cDNA cloning				
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)				
MEDLINE	99279253				
PUBMED	10349636				
REFERENCE					
AUTHORS	Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.				
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes				
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)				
MEDLINE	20499374				
PUBMED	11042159				



Db 422 CGTTACAAGATGATCTTCTACTTGGAGGGGGCTTCTGTGACTGTGGAGACACAGAAGCG 481  
QY 1164 TGAAGAACTGGCCCTTTTGTGTAAATCATGAACCTGGAAGAGCAGAGCTACTATATAAGAG 1223  
Db 482 TGAAGAACTGGCCCTTTTGTGTGATACAGAGCTGGAAGAGCAGAGCTACTATAAGAGAG 541  
QY 1224 AATTACAGCTGTCGGTTGAATGAAGAGGTAAATGTCCAAGCCAGGAGAAATATTTCCCTTCA 1283  
Db 542 AGCTTACATTTGCCCATTTGAATGAAGAGGTAAATGTCTCAAGCCAGGAGAAATATTTCCCTTCG 601  
QY 1284 GTGATAAAATATGCTGTAGAAATGACTATATGGAAGAGGAGAAAGAACTGCCCTCCCTGAA 1343  
Db 602 GTGATAAAATACATTTGAGAAATGACTATATGGAAGAGGAGAAAGAAATTTGCCCTCCCTGAA 661  
QY 1344 CTCAGATAGGGAGAGAAATGAAAGATACATTTGTGTCTCTTTTCAATGTATGAACACCAT 1403  
Db 662 CTCAGATAGGGAGAGAAATGAACGATACATTTGTGTCTCTTTTCAACGATGAGCACCAT 721  
QY 1404 TCATATGACCACTCATATACAGCCTACAAAGAGCTTTGACTGTGAGCTCGCAGAGGCC 1463  
Db 722 TCGTATGATCATGTGATCTACAGTCTGCAGAGAGCTCTAGATTGCGAGCTTGCAGAGGCA 781  
QY 1464 CAGTTTCATACCACTGCCATTTGACAAAGAGGGTCTGCGGCTGTTAAAGCGGGAGCTTAT 1523  
Db 782 CAGCTGCACAGCACTGCCATTCGACAAAGAGGGTCTGCGGCTGTTAAAGCGGGTGTAT 841  
QY 1524 GCTGCTTGCAGAGCAAGAGATATATAAGAGTCAATTCAGAAATGTCTCTCAACAT 1583  
Db 842 GCACTTGCAGAGCAAGAGGATATATAAGAGTCACTCAGAGAACGTCTCTCAGCAC 901  
QY 1584 CCACCTTCATAGAAATTTACACTCAGAGATTATGGCTCATCAGAAATTTGCTTTCCGT 1643  
Db 902 CCCTCCATGTGAAGTGTCTGCACCTCGTGTATGGCTCACCAGAAATTTCCGTCTCGGC 961  
QY 1644 CTTGGTTCCTGGATGAACAAATATGAGCTATTTCAAGTCACTTTAGCGAGATCTTTGC 1703  
Db 962 CTTGGCTCCTGGATGAACAAATATGAGCTATTTCAAGTCACTTTAGCAGATATTTGC 1021  
QY 1704 CAAGCATGCTTGAAGAGAACCTGACTCGGAGAAATCCCTGTCTCATAGCAGGTTAATG 1763  
Db 1022 CAGGCTCGCTGAGAGAACCTGGCTCTGAAATCCCTGTCTTATAGCAGACTAATG 1081  
QY 1764 CTTTGGATGCAAGGTTTATAAGGTGCCGTGAAGATCCTCATGAATTTGATCTTCAGC 1823  
Db 1082 CTTTGGATGCAAACTTTATAAGGTGCCGTGAAGATCCTCATGAATTTGATCTTTAGT 1141  
QY 1824 AGTTTTTTTATGAGATGGAATACAAAACTCTTTGCTATGAAATTTGTGAAGTATAT 1883  
Db 1142 AGTTTTTTTATGAGATGGAATACAAAACTCTTTGCTATGAAATTTGTGAAGTATAT 1201  
QY 1884 AAACAACTGCAGAAAGAAATATATCAGTGTATCATGACAGAGAGTATCTCTATAACTGCA 1943  
Db 1202 AAACAACTGCAGAAAGAGTATATCAGGACGACACCGAGAGAGCATCTCCATAACCGCC 1261  
QY 1944 CTTTCAAGTTCAGATGTTTACTGTCTTCTACTCTGGCTCGACATCTTATTTGAAGAGCAAT 2003  
Db 1262 CTGTCCGTGCAGATGCTCACCGTCCCGACCTTGGCCGGCATCTTATTTGAAGAGCAAT 1321  
QY 2004 GTTATCTCTGCTATTAAGTCTGCTAGAGAGTTTACTGAGTACTTTGAGCAGAGAAC 2063  
Db 1322 GTTATTTCTGCTATTACTGAAACGCTGTAGAGTTTACTGAAATTTTACTGAGAGAAC 1381  
QY 2064 AATAAATTCACCTCCAGGTTTATAGCAGGACAAATTTGGAGAGATATATGAGTAAATA 2123  
Db 1382 AATAAATTCACCTCCAGGTTTATAGCAGGACAAATTTGGAGAGATCTACCGAGTTATA 1441  
QY 2124 TGTGACCTAAAGTATATCTGTATCAGCAACCCACAATATGGACAGAAAGATTAAAGATG 2183  
Db 1442 TGTGACCTAAAGTATATCTGTATAGCAAGCCTGTCTATATGGACAGAACGATTAAAGAG 1501  
QY 2184 CAGTTCCCTTGAAGGTTTTCGATCTTTTGAAGATCTTTACCTGTATGCGAGGAATGGAA 2243  
Db 1502 CAGTTCCCTGGAAGGTTTCCGGTCTTTTCTGAAGATTTCTTACCTGTATGCGAGGAATGGAA 1561

RESULT 2  
BM472160  
LOCUS

BM472160

960 bp mRNA linear

EST 05-FEB-2002

QY 2244 GAAATCCGAAGACAGGTTTGGCAACACATTTGAAGTGGATCCTCTGATTTGGAGGCTGCCATT 2303  
Db 1562 GAAATCAGAAGACAAAGTTTGGCAACACATTTGAAGTGGACCTCTGACTGGGAGGCTGCCATC 1621  
QY 2304 GCTATACAGATGCAATTTGAAGAAATATTTTACTCATGTTCCAGAGTGGTGTGCTGTGAT 2363  
Db 1622 GCTATACAGATGCAATTAAGAAATATTTTCTCATGTTTCCAGAGTGGTGTGCTGTGAT 1681  
QY 2364 GAAGAACTCTTACTTGTGCTTTATAAAGAAATCTCAAAAGCTGTGATGAGGTGCATGACC 2423  
Db 1682 GAAGATCTCTTACTGTGGTGTCTTAAAGAAATGTCAAAAGCTGTAAAGAGTGCATGACATCA 1741  
QY 2424 AGTTTCTATATCTAGTAGCAAGACAGTAGTACAAATCGTGTGACATAGTTTGGAAACAAAG 2483  
Db 1742 AATTTCTATGTCAGTACCAAGACAGTAGTGCAATTTGTGCGTCAATAGTCTGGAACCAAA 1801  
QY 2484 TCCTACAGATATCTGAGGATCTTGTAGCATACTCTGCCACTCTCTAGGAGCCCTTGCT 2543  
Db 1802 TCCTACAAAGTCTCTGAGGACCTTGAAGCATACACCTGCCACTCTCTAGAACACTTGCT 1861  
QY 2544 GGTCTTCATGTACGTTTAAAGCAGGCTGGGTGCTGTTTCAAGACTGCATGAATTTGTGCT 2603  
Db 1862 GGTCTTCATGTACGTTTAAAGCAGACTAGTGCTATTTCAAGACTGCATGAATTTGTGCT 1921  
QY 2604 TTTGAGGACTTTTCAAGTAGAGTACTAGTGAATATCTCTTACGTTGTCTGTTGTTGCT 2663  
Db 1922 TTTGACAGCTTTCAAGTAGAGGCTCTGGTGGAGTACCCGCTGGCTGCTGCTGGTGGT 1981  
QY 2664 GCCAGGTTGTTGCTGAGATGGCGAAGAAATGGACTCTCTCTTAAATGACCAAGTGTGTT 2723  
Db 1982 GCTCAGGTTGTTGCTGAGATGGCGAAGAAACGGGCTCTCACTCATCAGCCAGGTTTTC 2041  
QY 2724 TATTACCAAGATGTTAAGTGCAGAGAAAGATGTATGATAAGATATCATGCTTCAG 2783  
Db 2042 TATTATCAAGATGTTAAGTGCAGGAGGAAATGTACGATAAGATATCATGCTTCAG 2101  
QY 2784 ATTGTGCTCTTTAATGATCCCAATAGTTCTTGTACTGTGACTTCCAGAGGTATGAA 2843  
Db 2102 ATTGAGCATCTATATGATCCCAACAGTTCTTGTACTGTACTTCCAGAGATATGAA 2161  
QY 2844 CTTGCCGAGGCTTTTAAACAGACCATATCTACAAAGACCAAGATTTGATTTAAACAATAT 2903  
Db 2162 CTTACTGATGCTTTTAAACAGACCATATCCACAAAGACCAAGATTTGATTTAAACAATAT 2221  
QY 2904 AATCACTAATAGAGAAATGCTTCAGGCTCTCATCTATATGCTGGTGGAGGCTGATGTA 2963  
Db 2222 AATCACTAATAGAGAAATGCTTCAGGCTCTCATCTATATGCTGGTGGAGGCTGATGTA 2281  
QY 2964 CTTGAGTGGGAAATGTGACCAAGAGAGGTCAAAATGAGAGAAATCATTTCACTTGCCT 3023  
Db 2282 CTTGAGTGGGAAATGTGACCAAGAGAGGTATATGAGAGAGATTTACTCACATCTT 2341  
QY 3024 TGCATTTGAACCCATGCCACAGTGCCTATGCCCCAAATTTACCTGAGAAATGAAATAAT 3083  
Db 2342 TGCATTTGAGCCCATGCCACAGTGCCTATGCCAGAAACCTACCTGAGAAACGAAATAAT 2401  
QY 3084 GAAACTGCTTAGAGAAATGTATATAACAAAGTGGCCACATTTAAAGAAACAGGTGTATCA 3143  
Db 2402 GAAACTGCTTAGAGAAATGTATATAACAAAGTGGCCACATTTAAAGAAACAGGTGTGCG 2461  
QY 3144 GGCCATGGAGTTTATGAATCAAAAGTCAATCACTGAAAGCTTCAATATGTACTTTTAT 3203  
Db 2462 GGCCATGGAGTTTATGAATCAAAAGTCAATCACTGAAAGCTTCAATATGTACTTTTAC 2521  
QY 3204 CATTACTCCAAACCCAGCATAGCAAGGCTGAACATATGC 3243  
Db 2522 CATTATTCTAAACACACATAGCAAGGCTGAACATATGC 2561

DEFINITION AGENCOURT\_6469668 NIH\_MGC\_92 Homo sapiens cDNA clone IMAGE:5576383  
5', mRNA sequence.  
ACCESSION BM472160  
VERSION BM472160.1 GI:18521202  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
NIH-MGC <http://mgc.nci.nih.gov/>.  
NATIONAL National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished  
AUTHORS Contact: Robert Strausberg, Ph.D.  
COMMENT Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
Tissue Procurement: ATCC  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: L1AM12328 row: c column: 08  
High quality sequence stop: 720.  
Location/Qualifiers  
1..960  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5576383"  
/tissue\_type="embryonal carcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_92"  
/note="Organ: testis; Vector: pCMV-SPORT6; Site: 1: NotI;  
Site: 2: SalI; Cloned unidirectionally; oligo-dr primed.  
Average insert size 2.5 kb. Library enriched for  
full-length clones and constructed by Life Technologies.  
Note: this is a NIH\_MGC Library."  
BASE COUNT 230 a 227 c 218 g 283 t 2 others  
ORIGIN  
Query Match 13.1%; Score 826.6; DB 12; Length 960;  
Best Local Similarity 96.1%; Pred. NO. 7.8e-160;  
Matches 890; Conservative 0; Mismatches 30; Indels 6; Gaps 4;  
QY 4924 TGTAGCATCCCATCCCTGTTGATGGGATGACCCCTGTTGATCGAGCCTCTTCAGTTA 4983  
DB 1 TGTAGCATCCCATCCCTGTTGATGGGATGACCCCTGTTGATCGAGCCTCTTCAGTTA 60  
QY 4984 GTTCTTCCATTAACACACCTTTATCTCTTCCATTTGATCACCATGGCACACATGCTTCAGA 5043  
DB 61 GTTCTTCCATTAACACACCTTTATCTCTTCCATTTGATCACCATGGCACACATGCTTCAGA 120  
QY 5044 TACTACTACAGTAGACACAGCGCTACCCCTGCTCAGGTTCAAGAAGACAGTGAAGAGG 5103  
DB 121 TACTACTACAGTAGACACAGCGCTACCCCTGCTCAGGTTCAAGAAGACAGTGAAGAGG 180  
QY 5104 CTCATTCGGCATCTCTTTCTTTCGAGAAATTTCTCAATATACAGTGGCTCCATGGGT 5163  
DB 181 CTCATTCGGCATCTCTTTCTTTCGAGAAATTTCTCAATATACAGTGGCTCCATGGGT 240  
QY 5164 GTGATATTCCTGGCTGGTATTTGGGGTCTCACTGAAGAATGGCATCCCGCTTATCTTC 5223  
DB 241 GTGATATTCCTGGCTGGTATTTGGGGTCTCACTGAAGAATGGCATCCCGCTTATCTTC 300  
QY 5224 GCTGTGCGCATGTTGTTTTCACATATTTACTTGGGGTAACTCCCGCTGAGGAATGCATA 5283  
DB 301 GCTGTGCGCATGTTGTTTTCACATATTTACTTGGGGTAACTCCCGCTGAGGAATGCATA 360  
QY 5284 CCAATTCGCAAGAGGAGTACAGTGCACCTCTGTAGTATCTATCTTTACCTACAAAT 5343  
DB 361 CCAATTCGCAAGAGGAGTACAGTGCACCTCTGTAGTATCTATCTTTACCTACAAAT 420

QY 5344 TGTTCCTGCTCTTCACAGGAATATTGGGATACTGTAAAGGCCCTTGCTCCAGAGGTGGTGTG 5403  
DB 421 TGTTCCTGCTCTTCACAGGAATATTGGGATACTGTAAAGGCCCTTGCTCCAGAGGTGGTGTG 480  
QY 5404 CAGATCCCTGCCCTTACTAAACTGTTTGAAGCAAAAAACACCGTGGTACCGTACCTAGAA 5463  
DB 481 CAGATCCCTGCCCTTACTAAACTGTTTGAAGCAAAAAACACCGTGGTACCGTACCTAGAA 540  
QY 5464 AAAGAAATAGTTTGATAGAGCTTCCCTGATGACTATAGCTGCCCTCCCTGAATCAAGCTTCTC 5523  
DB 541 AAAGAAATAGTTTGATAGAGCTTCCCTGATGACTATAGCTGCCCTCCCTGAATCAAGCTTCTC 600  
QY 5524 ATTTCAAGTGGCCACCGTCTGCAGATGATGAGGAAAGATCTGCTGCTCTGCTGCTTTCT 5583  
DB 601 ATTTCAAGTGGCCACCGTCTGCAGATGATGAGGAAAGATCTGCTGCTCTGCTGCTTTCT 660  
QY 5584 GTGGGGCTATATGTTCTCAGAACATTTGCTGCCAGGAAATTTGAAACGGGGAAGAGG 5643  
DB 661 GTGGGGCTATATGTTCTCAGAACATTTGCTGCCAGGAAATTTGAAACGGGGAAGAGG 720  
QY 5644 TTGGAGCTTGCATTTTTCACGCACCTTCACTGTGGAGCCGGAGTCTGCATTTTCTTAAAAA 5703  
DB 721 TTGGAGCTTGCATTTTTCACGCACCTTCACTGTGGAGCCGGAGTCTGCATTTTCTTAAAAA 780  
QY 5704 TCAGAGATGCGCAGAGTGG-TCCCTGGTTGAAGTAAAGCCAGAGGGTGT-CCCTATCCAGC 5761  
DB 781 TCAGAGATGCGCAGAGTGGTCTGTTGAAGTAAAGCCAGAGAGTGTGCGCTATCCAGC 840  
QY 5762 T-CCTTACTTGGATGAATATGGAGAA--ACAGACCTTGGCTTGAAGGGGCAACCC 5817  
DB 841 TCCCTTACTTGGGATGATATGAAAAACAGAACCTTGGGCTTGAAGGGGCAACCC 900  
QY 5818 TTCAATTTATCTCGTGAGCGGTATCGG 5843  
DB 901 CCCTTTCATTTATCTCGTGAGCGG 926  
RESULT 3  
BC044903  
LOCUS 3641 bp mRNA linear HTC 10-FEB-2003  
DEFINITION Mus musculus, Similar to ubiquitin ligase E3 alpha-II, clone  
IMAGE:3493115, mRNA.  
ACCESSION BC044903  
VERSION BC044903.1 GI:28279967  
KEYWORDS HTC.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
Strausberg,R.  
Direct Submission  
Submitted (23-JAN-2003) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
Tissue Procurement: Gilbert Smith, Ph.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Baylor College of Medicine Human Genome  
Sequencing Center  
Center code: BCM-HGSC  
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>  
Contact: [ang@bcm.tmc.edu](mailto:ang@bcm.tmc.edu)  
Gunnaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulseged, H.,  
Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,  
A.N., Gibbs, R.A.  
Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>



Series: IRAK Plate: 40 Row: b Column: 10  
This clone has the following problem: frame shifted.

FEATURES  
source

Location/Qualifiers  
1.3641  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="CZECH II"  
/db\_xref="taxon:10090"  
/clone="IMAGE:3493115"  
/tissue\_type="Mammary tumor metastasized to lung. Tumor  
arose spontaneously from a senescent normal mammary  
(clonal) outgrowth infected with the virus MMTV."  
/clone\_lib="NCI\_CGAP\_Lu29"  
/lab\_host="DH10B"  
/note="Vector: pCMV-SPORT6"

BASE COUNT 989 a 899 c 906 g 847 t  
ORIGIN

Query Match 12.8%; Score 806.4; DB 11; Length 3641;  
Best Local Similarity 54.7%; Pred. No. 1.3e-155;  
Matches 1934; Conservative 0; Mismatches 1516; Indels 88; Gaps 13;

2460 TGTGGACATAGTTTGGAAACAAAGTCTACAGAGTATCTGAGGATCTTGAAGCATACAT 2519  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
22 TGTGGACACTCGGTGGAAACCAACAGATACCTGTTTCCCAAGAAAGTTAGCATTCAC 81

2520 CTGCCACTCTAGGACCCCTTGCTGCTTCATGCTAGCTTTAAAGCAGGCTGGGTGCTTT 2579  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
82 CTCCCAATTTCTGCTGCTGAGGTTTGCATGATTTGTAACCAAAAGTGAAGTGCCA 141

2580 TCAAGACTGCATGAATTTGCTTTTGGAGACTTTTCAAGTGAAGTACTAGTGAATAT 2639  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
142 TATAAATTTCCAGAGCTCTACCTCTAAGTGAAGTGAAGCCACCCATGTTGATAGCAT 201

2640 CTTTACGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2699  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
202 CTTTACGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 261

2700 CTGCTCTTATAGCAGGCTGTTTATACCAAGATCTTAAGTGCAGAGAGAAATGAT 2759  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
262 TTCTCTCTAGTAATCAGATCTATCTACCAATGTAATGTAATGTAATGTAATGTAATGTA 321

2760 GATPAAGATATCATCTATCTAGATTTGGTGCATCTTTAATGATCCCAATTAAGTCTTGT 2819  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
322 GACAAGGACATAGTATGCTTTCAGACAGGTTCTCCATGATGATGATGATGATGATGATGAT 381

2820 TTTACTGGTACTTCAGAGGTATGAACCTTCCGAGGCTTTTAAAGACCATATCTACAAA - 2878  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
382 ATGATCATGCTACGCGCTTGAACCTCTATCAGCTCTTTCAGCAGCCTGACTATGGGAAG 441

2879 -----AGACCAGGATTTGATTAACCAATATATACACTAATAGAA 2918  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

442 AGATTCAGTTCTGAGTTACCCATAAGGACGTCGTTTCCAGAGAAACACTCTCATCGAA 501

2919 GAAATGCTTCAGGTCCTCATCTATATTTGGGTGAGCGTTATGTACCTCGAGTGGGAAAT 2978  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

502 GAGATGCTTACCTCATCATCTGCTTGTGGGAAAGATTCACCCCTGGGGTTGGACAG 561

2979 GTGACCAAGAGAGGTCACATAGAGAAATCATTCACCTTGCTTTGCTATGAACCCCATG 3038  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

562 GTGGCTGCCACAGATGAATCAAGAGGAGATTTATCCATCAGCTGAGCATCAAAACCTATG 621

3039 CCACACAGTCCCATTTGCCAAAATTTACCTGAGAAATGAAATATGAACTGGCTTAGAG 3098  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

622 GCTCAGAGGAGCTGGTGAAGTCTGCTGCTGAGATGAGAACAGGAGACCGGATGGAG 681

3099 AATGTCTATAAACAAAGTGGCCACATTTAAGAAACCGAGTGTATCAGGCCATGGAGTTAT 3158  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

682 AGCGTTCAGGTCGCGGACACATTTCAAGAAACCTGGCCTCAGACGACGAGCATGAT 741

3159 GAACATAAAGATGAATCACTGAAGACTTTCAATATGTACTTTTATCTACTCAAAACC 3218  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

742 GAGCTGAAGCCAGAGTGTGCCAAGAGTTCAACCTGTATTTTATCATTTTCTCCAGGCA 801

QY 3219 CAGCATAGCAAGCTGAACATATGACAGAAAGAGAGAAACAAAGAAATGAA 3278  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 802 GAGCAGTCCAAGGACAGGAAGCTCAGCGAAATTTGAAAGAGAGAAATAAGAGATACA 861

QY 3279 GCATTGGCGCCACCAACCTCCTGTAATTTCTGCCCTGCTTTTACGAAAGTGAATTAACCTT 3338  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 862 GCATCTCCCTCCCTCCGGCTTTGGCACCGTTCGCGCTTTGTTCCGGAGTCTGGTTAACATC 921

QY 3339 CTCAACTGTGATATCATGATGATACATCTCAGGACCGTATTTTGGCGGGCAATAGACACA 3398  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 922 TTGCAGTGTGACGTCATGCTGTACATCATGCGGAACCATCTGCGAGTGGCTGTAGAGCAT 981

QY 3399 GATTTAACTTTGTGACCGGAGGATGCTCCAAATGGCTTTTTCATATTTCTGGCATTTGGT 3458  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 982 CACGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1041

QY 3459 TTACTAGAGAGAGCAACAGCTTCAAAAAGCTCTCGAGAGAAAGT---AATTTGAC 3515  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 1042 CTCAGAGAGAGAGACCACTTTGGAGAAATGCGTGGAAAGGCTATGCGAGACCTTCAACC 1101

QY 3516 TTTTATCATTAAGGCTTCAAGATTTGGGAAGTTCAG-----CCATGATATACAAATG 3566  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 1102 TTCACAGAGAGATTTCAAAAGCCTGCTGATGCACACATAACTCCCCGAGCATCTTAGCT 1161

QY 3567 CTTTGGAAAAACTCAAGGAATTTCCCAAGTTAGAAAGGCGAGAGGACATGATAAGCTGG 3626  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 1162 ATGCTGGAGACCTTCAGAACCCCTCCCTCCCTGGAAGCCCAAGGACATGATCAGGTGG 1221

QY 3627 ATACTTCAGATGTTTGCACACAGTGAAGCATTAAAGAGAAAAATTTGTTTAAATTTAGCA 3686  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 1222 TTGCTAAAGATTTTAAATGCAATTAAGAAAGATAAGAGA-----GTGTTTCATCCAGCAGC 1275

QY 3687 ACCACATCAGGATTCGGAATCTATTAAAGATGATGAGATTTACTCATGATAAAGAAAAAGCA 3746  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 1276 CCGCTGGCGAGGCGGAGGGAACCAATAATGGAGAGAGCTCAAGAGACAAAGGACAAAGCA 1335

QY 3747 GAACCAAAAGAAAAAGCTGAAGCTGCTAGCTACATCGCCAGAGAAATCATGGCTCAGATG 3806  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 1336 GAGAGAAAGAAAGAACCCAGATGCGCAGACTGCGCGGAGAGATCATGGCCAGATG 1395

QY 3807 TCTGCTTTACAGAAAAAATTCATTTGAAACTCATAAACTCATGTATGACATATACATACAGAA 3866  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 1396 TCTGAGATGCGAGCGCACTTTCATTTGACGAAACAAAGAGCTCTTCCAGCAGACCTTAGAG 1455

QY 3867 ATGCTGGGAAAGAGATTTCCATTTATGGAGAGAGAGACACCCAGCTGACTGACTAC 3926  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 1456 CTGGACACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1509

QY 3927 TCTAGAATTTGCTTGGGTCCTAAACGGGTCCATCTGTTACTGAAAAGGAGGTGCTGACG 3986  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

Db 1510 GCTCTTACAGACTGGGCCACACAGACAGAGTCCCTGAAACCGAGACATGTTGTCACC 1569

QY 3987 TGCACTCTTTGCAAGAAAGACAGAGGTGAAATAGAAAATATGCCATGTTATTCG 4046  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

Db 1570 TGTATATTATGTCAGAGAGAGCAAGAGGTGACTGTGGGAGAGAGGCGATGCTTGGCA 1629

QY 4047 GCCTGTGTCAGAAATCTACTGCTTTAACCCAGACAGAGGGAACCCATAGAACTCTCA 4106  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

Db 1630 GCATTTGTTTCAGAGGTCAACGGTCTGTCTCAAAAGACAGACGAAACCACTCGCGGACCCA 1689

QY 4107 GGAGAACCCCTAGACCCACTTTTCATGGATCCAGACTTGGCATATGGAATTTATACAGGA 4166  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

Db 1690 GAAAAA---TATGATTCATTTATTCATGACCCCGATCTCTCTGTTGGGACACACATGGC 1746

QY 4167 AGCTGTGCTGATTAATGACGACAGTGTGCTGGCAGAACTATTTTCAAGCTGTAC----- 4221  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

Db 1747 AGCTGTGGCAGCTTATGATGCCCATTTGTTGGCAAGAGTATTTGATTCGTTCAAGCC 1806

QY 4222 ----AGCTGAGCTCTCAGCAGCGCATTTTCATGTTGACCTT---TTTGTGGAAGTGA 4274  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

Db 1807 AAGGAGCAGCGAAGCAGCAGCGGCTGCGCTTGCACACTAGCTACCATGTAGAGAAATGCG 1866

Qy		4275	GAAATATTCTTGGCCCTCTTTGCCAATCTCTGTGCAATCTGTGCAGTAACCTTCATCCTTG 4339
Dd		1867	GAGTTCTCTCGCCGCTCTGTAGTGCTTGAAGAACACGGTGATCCCC-----CTGCTG 1920
Qy		4335	CACCCTCAAAGATAACAAGTGCAGATGCGTCTTGGCTCAACTTTTGACCCCTGGCA 4394
Dd		1921	CTTCCTCCCAGGAGCATCCTCGGAGAGGTTAAAATTTTFCAGACCAACACGATCTGGCA 1980
Qy		4395	CGGTGTATCACAGACTGTCTTGGCCAGAAATATCAGGTTATATAATTAAGACATGCTAAAGGA 4454
Dd		1981	CAGTGGAGAGAGCATTAACGACAGCAATAAAGGTGGTCCAGATGCTGAGGAGAAGACAC 2040
Qy		4455	GAAACCCAATTCCTATTTTTCTTAATCAAGGAATGGGAGATTTCTACTTTGGAGTTCCAT 4514
Dd		2041	AGTGCTGCTGACACGCTCTTCTTCAGAGGACACAGAAGCCATGAATAATAATACCGATCCCC 2100
Qy		4515	TCCATCTCTGAGTTTTGGCGTTGAGTCTTTCATTAATAATTAATTAATTAATTAATTAATTA 4574
Dd		2101	GAAGGCTTCAGGCGCTGATTTTTATCCTTAGAACCCCATATTTCTGATAGCATAAAGAANAATG 2160
Qy		4575	GTTATTCTCTTGGCCACAACAATTTATAGAAATTTGGATTGGAAGTGGCACCTGATGAAGG 4634
Dd		2161	TTAACGACATTTGGAAGGCTGCTTACAAGTGGGACTGAAGTTTATCTTAATGAAGGT 2220
Qy		4635	GATCCTCGAGTCCCCATGCTGACCTGGAGCACCTGGCGTTTTCATCTPCCAGGCAATTGAA 4694
Dd		2221	GACCCCGTGTGCCATCTGTGCTGGGGACCTGTGCATACACCATCCAGAGCATAGAA 2280
Qy		4695	AATCTATTGGGAGATGAAGAAAAACCTCTCTTTGGAGCACTTCAAAATAGGCAGCATTAAT 4754
Dd		2281	AGAAITTTGAGTGATGAGGAGAGCCCTGTTTTTGACCTCTGCTTGTAGACTGGACGAC 2340
Qy		4755	GGTCTAAAGCATTAATPGCAGTTTGCAGTTGTCACAGAGGATTAACCTGCTCCTCAGGCTTG 4814
Dd		2341	TGTCAGGCTGTTAAACAGGTTTGCAGCAGCACAATTGGACAGTGGCGTTACTTCTGCTG 2400
Qy		4815	ATACAGAAACATCTGGTTCGTCTTAATCAGTTGTTCCTAACATFAAAATCAGAAGAT 4874
Dd		2401	GTACAGGACACTTCTGTAACACTCTTTGTCATCTTGGTGCCTAGTGACAGCTATGAAGC 2460
Qy		4875	ACACATGCTCTGCTATAGATCTGTTTTCATGTTTGGTGGGTGCTGTGTTAGCATTC 4934
Dd		2461	CTCCGCTGCATPACATAGACATTCGACATGTTTCACCTGCTGTGGCGCCCTGGTCTCGCTTC 2520
Qy		4935	CCATCTTGTATTGGGATGACCCCTGTTGATCTGCAGCTCTTCTCAGTTAGTTCTTCTCTAT 4994
Dd		2521	CCAGCTCTGAGTG-----TCAGGATTTTTCAGGAAGCACCCCTGGCCACT 2565
Qy		4995	AACACCTTTATCTCTCCATTTTATCACCATGGCACACATGCTTCAGATACTACTTACA 5054
Dd		2566	GGAGACCTGCACATCTTCCACCTGGTTACCATGGCACACATCGTACAGACTTACTTATCC 2625
Qy		5055	GTAGACAC---AGGCCCTACCCCTTGTCTCAGGTTCAAGAAGACAGTGAAGAGGCTCATTC 5111
Dd		2626	TCATGTACAGAAGAGAATGGCATGGATFCAAGAGAATCCCATCGGGGAAGAGAACCTGGCC 2685
Qy		5112	GCATCTTCTTTTGCAGAAATTTCTCAATATACAGTGGCTCCATTTGGGTGTGATATT 5171
Dd		2686	ATTCTCTCTTTCGACAAAACACTTCACCATGATATCTGGAAGTGCC-JTGAAGAAGGCCCC 2744
Qy		5172	CCTGGCTGGTATTTGTGGGTCTCACTGAAGAATGGCATCCACCCTTATCTTCGCTGTGCT 5231
Dd		2745	TCCGGCTGGCACCTGTGGAGGCGCCGCGGCCCATCATGCTTTTCCCTCAAGTGTCT 2804
Qy		5232	GCATTTGTTTTCCACTATTTTACTTTGGGGTAACCTCCGCGCTGAGGAACGTGCATATTCCT 5291
Dd		2805	GCTTTGTTTTTCCACTATTTAAATGGAGTCCCGGCCCTCCAGACCTTCAAGT----TTCT 2861
Qy		5292	GCAGAAGGAGTACAGTGCACCTGTGTAGCTATCTATCTTTTACCTACAAATTTGTTCCCTG 5351
Dd		2862	GGAACAAGCCATTTTGAACACTTATGTATACCTTTTCCCTTACCAACCAACCTCATATTCAC 2921
Qy		5352	CTCTTCCAGGAATATTGGGATCTGTGAAGGCCCTTGTCTCCAGAGGTTGGTGTGCAGATPCCT 5411

Db	2922	CTTTTTCAGAAAAACAGTGCATCATGAACCTCCCTGATTGAAAGTTGGTGCCAGAACAGT	2981
QY	5412	GCCTTACTAACTGCTTTTGAAGCAAAAAACACCGTGCAGTACCGTACAGGTAACCAAGAAAGAAAT	5471
Db	2982	GAAGTTAAACGGTATCTAAATGGCGAGAGAGCGATAAGCTACCCAGCAGGAGCTAAC	3041
QY	5472	AGTTTGTATAGAGCTTCCTGATGACTATAGCTGCCTCCTGAATCAAGCTTCTCATTTTCAGG	5531
Db	3042	AAACTGATAGACCTTCAGAGAGATTACAGCAGCCTCATTAACCAAGCATCCAATTTCTCG	3101
QY	5532	TGCCACGGCTGCAGATGATGAGCGAAAGCATCCTGCTCTGCTGCCTTTTCTGTGTGGGCT	5591
Db	3102	TGCCCAAAATCAGGTGGCGACAAGACGACAGAGCTCCTACTCTGTGCTCGTGTGTGGAGT	3161
QY	5592	ATACTATGTCTCAGACATTTGCTGCCAGGAATTTGTGAACGGGAAGAGTTGGAGCT	5651
Db	3162	CTCCTCTGCTCTCAGAGTTACTGCTGCCAAGCTGAGCTGGAGGTGAGGACGTCGGAGCC	3221
QY	5652	TGCATTTTTCACGCACACTCACTGTGGAGCGCGAGTCTGCATTTTTCCTTAAAAATCAGAGAA	5711
Db	3222	TGCACAGCACACCTACTCCTGTGGCTCCGGGCGCGCATCTTCCTGAGATTCGGGAA	3281
QY	5712	TGCCGAGTGGPCTGTGGTGAAGGTAAAGCCAGAGGCTGTGCCTATCCAGCTTCCTTACTTG	5771
Db	3282	TGTGAGGTGCTATTTTGTAGTGGCGAAACCAAGAGATGTTTTATTCTCCTCTTACCTT	3341
QY	5772	GATCAATATGAGAAACAGACCCCTGGCTCTGAAGAGGGCAACCCCTTCATTTATCTCTCGT	5831
Db	3342	GACGACTTGGGGAGACCGACGAGGACTCAGACGAGGAATCTTCATATTATGCGAA	3401
QY	5832	GAGCGGTATCGGAAGCTCCATTTTGGTCTGGCAACAACACTGCATATTAGAGAGATTGCT	5891
Db	3402	GAGCGGTTTCGAAAGATCCAGAACTCTGGCAGCAGCATAGTATCAGAGGAGATCGGA	3461
QY	5892	AGGAGCCAGAGACATAATCAGATGTTATTTGGATTCAACTGGCGAGTTACTGTGAGCTC	5949
Db	3462	CACGCGCAGGAGGCTAACCCAGACCCCTGTCGGAATTGACTGGCAGCATTTATAATCGC	3519
RESULT 4	BQ233617	BQ233617	
LOCUS	AGENCOURT_7575596 NIH_MGC_72 Homo sapiens	cdna linear	EST 02-MAY-2002
DEFINITION	5', mRNA sequence.		
ACCESSION	BQ233617		
VERSION	BQ233617.1	GI:20415017	
KEYWORDS	EST.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	1 (bases 1 to 865)		
JOURNAL	NIH-MGC http://www.ncbi.nlm.nih.gov/.		
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC)		
	Unpublished		
	Contact: Robert Strausberg, Ph.D.		
	Email: cgapbs-r@mail.nih.gov		
	Tissue Procurement: ATCC/DCTD/DTF		
	cdna Library Preparation: Life Technologies, Inc.		
	CNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)		
	DNA Sequencing by: Agencourt Bioscience Corporation		
	Clone distribution: MGC clone distribution information can be		
	found through the I.M.A.G.E. Consortium/LLNL at:		
	http://image.llnl.gov		
	Plate: LLNL13294 row: 1 column: 04		
	High quality sequence stop: 720.		
FEATURES	Location/Qualifiers		
source	1..865		
	/organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:6047379"		





ORGANISM Homo sapiens  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 712)  
 Ebert, L., Heil, O., Hennig, S., Neubert, P., Partsch, E., Peters, M.,  
 Radelof, U., Schneider, D. and Korn, B.  
 Human Unigeneset - RZPD3  
 Unpublished  
 Contact: Ina Rofls  
 RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH  
 Im Neuenheimer Feld 580, D-69120 Heidelberg, Germany  
 RZPD; IMAGp998F081824.  
 RZPDLIB; I.M.A.G.E. cDNA Clone Collection;  
 Human Unigeneset - RZPD3 (RZPDLIB No.972)  
 http://www.rzpd.de/CloneCards/cgi-  
 bin/showlib.pl.cgi?response=libNo=972 Contact: Ina Rofls  
 RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH  
 Heubnerweg 6, D-14059 Berlin, Germany  
 Tel: +49 30 32639 101  
 Fax: +49 30 32639 111  
 www.rzpd.de  
 This clone is available royalty-free from RZPD;  
 contact RZPD (clone@rzpd.de) for further information. Seq primer:  
 M13r, Primer sequence: TTTCACACAGGAACAGCTATGAC.  
 Location/Qualifiers  
 1..712  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGp998F081824 ; IMAGE:742663"  
 /sex="male"  
 /lab\_host="DH10B"  
 /clone\_lib="Soares\_testis\_NHT"  
 /notes="Vector: pT7T3D-Pac (Pharmacia) with a modified  
 polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA  
 was prepared from mRNA obtained from Clontech Laboratories  
 Inc., and primed with a Not I - oligo(dT) primer [5',  
 TGTTCACCATCTGAAGTGGAGCGCGCCCAATTTTTTTTTTTT 3'].  
 Double-stranded cDNA was ligated to Eco RI adaptors  
 (Pharmacia), digested with Not I and cloned into the Not I  
 and Eco RI sites of the modified pT7T3 vector. Library  
 went through one round of normalization to Cot5, and was  
 constructed by Bento Soares and M. Fatima Bonaldo."  
 BASE COUNT 174 a 164 c 173 g 201 t  
 ORIGIN  
 Query Match 11.3%; Score 710.4; DB 13; Length 712;  
 Best Local Similarity 99.9%; Pred. No. 7.4e-136;  
 Matches 711; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 5173 CTGGCTGGTATTTGGTGGTCTCACTGAAGAATGGCATCCCTTATCTTCGCTGCTG 5232  
 Db 1 CTGGCTGGTATTTGGTGGTCTCACTGAAGAATGGCATCCCTTATCTTCGCTGCTG 60  
 QY 5233 CATTTGTTTCCACTATTTACTTGGGGTAACCTCCGCTGAGGAACATGACCAATCTG 5292  
 Db 61 CATTTGTTTCCACTATTTACTTGGGGTAACCTCCGCTGAGGAACATGACCAATCTG 120  
 QY 5293 CAGAAGGAGTACAGTGCACCTCTGAGTATCTATCTTTACCTACAAATTTGCTCTGC 5352  
 Db 121 CAGAAGGAGTACAGTGCACCTCTGAGTATCTATCTTTACCTGCAAAATTTGCTCTGC 180  
 QY 5353 TCTTCAGGAATATGGGACTGTGAAGGCCCTTGCTCCAGAGCTGGTGTGACAGTCTG 5412  
 Db 181 TCTTCAGGAATATGGGACTGTGAAGGCCCTTGCTCCAGAGCTGGTGTGACAGTCTG 240  
 QY 5413 CCTTACTAACTGTTTGAAGCAAAAAACACCGTGGTGCAGTACCCTAGAAAAAAGAAATA 5472  
 Db 241 CCTTACTAACTGTTTGAAGCAAAAAACACCGTGGTGCAGTACCCTAGAAAAAAGAAATA 300  
 QY 5473 GTTTGATAGAGCTTCTGTAGTACTATAGTGCCTCTGATCAAGCTTCTCATTTCAAGT 5532  
 Db 301 GTTTGATAGAGCTTCTGTAGTACTATAGTGCCTCTGATCAAGCTTCTCATTTCAAGT 360

QY 5533 GCCACAGGTCTGCAGATGATGAGCGAAAGCATCTCTCTCTGCGCTTTTCTGTGGGCTA 5592  
 Db 361 GCCACAGGTCTGCAGATGATGAGCGAAAGCATCTCTCTCTGCGCTTTTCTGTGGGCTA 420  
 QY 5593 TACTATGTTCTCAGAACATTTCTGCCAGGAATGTGAACGGGAGAGGTTGGAGCTT 5652  
 Db 421 TACTATGTTCTCAGAACATTTCTGCCAGGAATGTGAACGGGAGAGGTTGGAGCTT 480  
 QY 5653 GCATTTTTCAGCACTTCACTGTGGAGCGCGAGTCTGCAATTTTCCCTAAAAATCAGAGAAT 5712  
 Db 481 GCATTTTTCAGCACTTCACTGTGGAGCGCGAGTCTGCAATTTTCCCTAAAAATCAGAGAAT 540  
 QY 5713 GCCAGTGGTCTCTGTTGAAGTAAAGCCAGAGGCTGCGCTATCCAGCTCCTTACTTGG 5772  
 Db 541 GCCAGTGGTCTCTGTTGAAGTAAAGCCAGAGGCTGCGCTATCCAGCTCCTTACTTGG 600  
 QY 5773 ATGAATATGGAGAACACAGCCCTGCTGAACAGGGGCAACCCCTTCAATTTATCTCCTG 5832  
 Db 601 ATGAATATGGAGAACACAGCCCTGCTGAACAGGGGCAACCCCTTCAATTTATCTCCTG 660  
 QY 5833 AGCGTATCGGAAGCTCCATTTGGTCTGGCAACACACTGCATTATAGAAGA 5884  
 Db 661 AGCGTATCGGAAGCTCCATTTGGTCTGGCAACACACTGCATTATAGAAGA 712  
 RESULT 7  
 AK083320  
 LOCUS  
 DEFINITION Mus musculus 2 days neonate thymus thymic cells cDNA, RIKEN  
 full-length enriched library, clone:C920004H05 product:similar to  
 (A PROTEIN SIMILAR TO UBIQUITIN-PROTEIN LIGASE E3 COMPONENT  
 N-RECOGNIN (UBIQUITIN-PROTEIN LIGASE E3-ALPHA)) (FRAGMENT) [Homo  
 sapiens], full insert sequence.  
 ACCESSION AK083320  
 VERSION AK083320.1 GI:26350448  
 KEYWORDS HTC; CAP trapper.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 1  
 Carninci, P. and Hayashizaki, Y.  
 High-efficiency full-length cDNA cloning  
 Meth. Enzymol. 303, 19-44 (1999)  
 99279253  
 PUBMED 10349636  
 2  
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
 Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
 Normalization and subtraction of cap-trapper-selected cDNAs to  
 prepare full-length cDNA libraries for rapid discovery of new genes  
 Genome Res. 10 (10), 1617-1630 (2000)  
 20499374  
 PUBMED 11042159  
 3  
 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,  
 Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M.,  
 Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,  
 Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,  
 Fujiwara, S., Inoue, K., Todawa, Y., Izawa, M., Ohara, E., Watahiki, M.,  
 Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J.,  
 Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
 RIKEN integrated sequence analysis (RISA) system--384-format  
 sequencing pipeline with 384 multicapillary sequencer  
 Genome Res. 10 (11), 1757-1771 (2000)  
 20530913  
 PUBMED 11076861  
 4  
 Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,  
 Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S.,  
 Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I.,  
 Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R.,

Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nikaado, I., Pesole, G., Quackenbush, J., Schriml, L.M., Staubli, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M.F., Brownstein, M.J., Bult, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, M., Hume, D.A., Kamiya, M., Lee, N.H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P., Ring, B., Rinchard, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K.F., Suzuki, H., Toyooka, K., Wang, K.H., Weitz, C., Whittaker, C., Wilming, L., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kotsuki, S. and Hayashizaki, Y.

Functional annotation of a full-length mouse cDNA collection  
Nature 409 (6821), 685-690 (2001)  
21085660  
11217851

**TITLE**  
The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.

**JOURNAL**  
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

**REFERENCE**  
Nature 420, 563-573 (2002)

**AUTHORS**  
6 (bases 1 to 3627)

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurahara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Soabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takai-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.

**TITLE**  
Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp, URL: <http://genome.gsc.riken.go.jp/>, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

**COMMENT**  
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Tissues were provided by Dr. John Todd (Dept. of Medical Genetics Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome Trust/MRC building Addenbrookes Hospital Cambridge) whose assistance we gratefully acknowledge.

Please visit our web site for further details.  
URL: <http://genome.gsc.riken.go.jp/>  
URL: <http://fantom.gsc.riken.go.jp/>.  
Location/Qualifiers

**FEATURES**  
source  
1..3627  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="FANTOM\_DB:C920004H05"  
/db\_xref="taxon:10090"  
/clone="C920004H05"  
/cell\_type="thymic cells"  
/tissue\_type="thymus"  
/clone.lib="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="2 days neonate"  
1044..>3626  
/note="unnamed protein product; putative  
similar to (A PROTEIN SIMILAR TO UBIQUITIN-PROTEIN LIGASE

E3 COMPONENT N-RECOGNIN (UBIQUITIN-PROTEIN LIGASE E3-ALPHA) (FRAGMENT) [Homo sapiens] (SPTR|Q9H578, evidence: FASTY, 78.8%id, 100%length, match=749), /codon\_start=1  
/protein\_id="BAC38864.1"  
/db\_xref="GI:26350449"  
/translation="MFRYAVDILITWEKSELPEDEVAEKSDDTYCYMLNDEVHTYEQ VITLQKAVNCTQKEAIGFATTVDRGRSVRYGDFQYCDQAKTVIVRNSRQTKPLK VOVMHSSVAHQNFGLKALSWLGSVIGYSDGLRRILCOVGLQPGDGSLSVDRML NDSKMLKWSAGVYHQLFMSLLMDLYKKLFAIRFAKNYERLQDYVTDHDFEVSVA DSVQITVPSLARMLLTEENLMTVIIKAFMDHLKRDQAGREFQRYTALQAFKFR VLQILDLKYLKSPTEWDELQKFLQGFADLELLKCMQGMDPITRQVGOHIEME PWERAFTLOMLKTHVISWQDWCALDEKVLIEAYKKCLAVLQCHGGFTDGGQPITL SICGHSVETIRYCSQEKVSIHLPIISRLNLAGLHVLSKSEVAFKPELLPSLSPPM LIEHLRCLVLCQAQVHAGMRNFSLVNIYYHNKCRREMDKDIVMLQGVSMML LLEHPLRMLMSRFDYQLFSTYIKPMASSELVSLVQONNTLIEMLYLIIMLVGE RNPVGOVAATDEIKREITIHOLSIKPMASSELVSLVQONNTLIEMLYLIIMLVGE RNPVGOVAATDEIKREITIHOLSIKPMASSELVSLVQONNTLIEMLYLIIMLVGE KPCLTGRMYELKPECAKBNLYFYHFSRAEOSKARENOKLREKEDPALPPALP PFCPLPASLVNIIQCDVMYIIMGITLOWAVEHGHSAWSESLORVLHLMGALQEEKH HLENAVEGHVQTFITFKISKDGDAPHNSPILAMLETQNLQNSLEAHKMDIRWLKM FNAIKKIRECSSSPVAEAGTITMESSRDKDKAERKAEIARLRREKTAQMFEMQ RHFD"

BASE COUNT	943 a	849 c	1012 g	823 t
ORIGIN				
Query Match	10.9%	Score 687.8;	DB 11;	Length 3627;
Best Local Similarity	56.0%;	Pred. No. 4.2e-131;		
Matches 1472;	Conservative	0;	Mismatches 1107;	Indels 51;
Gaps	7;			
QY	1245	GAAGAGGTAATGTCTCAAGCCAGGAGAAATATTTTCCTTCAGTGATAAAATATGCGTAGAA	1304	
DB	1005	GAAGATGTGATCGCCAGAGACTTACAACATTTTTCCTTATTTGTCGATATGACGATAGAT	1064	
QY	1305	ATGACTATATGGAGAGGAGAAAGAACTGCCCTCCCTGAACCTCCAGATAGAGGAGAAAT	1364	
DB	1065	ATACTGACCTGGGAAAAAGAAAGTGAATTCCTTGAAGACTTAGAAGTGGCAGAGAGAGT	1124	
QY	1365	GAAGATACATATTTGTCCTTTTCAATGATGAACACACATTCATATGACACGCTCATATAC	1424	
DB	1125	GACACCTACTACTGCTGCTGTTTAAATGATGAGGTTTACACCTTATGAGCAAGTCATTTAT	1184	
QY	1425	AGCTTACAAGAGCTCTTGACTGTGAGTCGCGAGAGCCCGAGTTCATACCTACCTGCGCAT	1484	
DB	1185	ACCTTCAGAAAGCTGTGAAGCTGACACAGAAAGCCATTCGCTTTGCAAGTACAGTT	1244	
QY	1485	GACAAAGAGGTCGTCGGCTGTAAAAGCGGAGCTTATGCTGCTGCCAGGAGCAAG	1544	
DB	1245	GATCGAGATGGCGTAGTCTGTCGGATATGGAGATTCAGTACTGTGATCAAGCAAG	1304	
QY	1545	GAAGATATAAGAGTCATTCAGAAATGTCCTCAACATCCACTTTCATAGTAGATATTA	1604	
DB	1305	ACAGTCATTTGTAGGAACACACAGCAGACAGACC---AAGCCGCTCAAAGTTCAGTATG	1361	
QY	1605	CACCTCAGAGATATGCTCATCAGAAATTTGCTTGGCTTGGTTCCTGATGAGCAAGAA	1664	
DB	1362	CACCTCCTCGCTGCTGCTCATCAGAAATTTGTTGTTGAAAGCTCTGCTGCTGGGAGT	1421	
QY	1665	ATTATCAGCTATTCAAGTACTTTTAGCAGATCTTTTGGCAAGCATGCCCTTAGAGAGAA	1724	
DB	1422	GTATTGGATACCTCAGTGGCCTTCGACAGGATTTTGTCTCAAGTTCGATACAGAGAGT	1481	
QY	1725	CCTGACTCGGAGAAATCCCTGTCTCATTAAGCAGGTTAATGCTTTGGGATGCAAGCTTTAT	1784	
DB	1482	CCAGATGGCGCAAAACTCTTCTCTGGTCGACAGACTGATGCTTAATGATTCCAAATATGG	1541	
QY	1785	AAAGGTGCCCGTAGAATCTTTCATGAATTCATCTTCAGCAGTTTTTTTTATGGAGATGAA	1844	
DB	1542	AAAGGGCTAGGAGTGTGTATCATCAGCTTGTTCATGAGCAGCTGCTCATGACCTCAAG	1601	
QY	1845	TACAAAAAACTCTTTGCTATGGAATTTGTCAAGTATTATAAACAACGTCGAGAGAAATAT	1904	
DB	1602	TATAAGAGCTGTTCGGCTTCGATTTGCTTAAAACTATGAGCGCTTTCGACAGAGATAT	1661	

CDS

QY	1905	ATCAGTGAATGATCATGACAGAAAGTATCTCTATAACTGCACATTTTCAGTTTCAGATGTTTACT	1961
Db	1662	GTGACAGATGACCCAGCAGAGAGATTTTTCAGTCGCAGACCTCTCGGTTTCAGATATTCACA	1721
QY	1965	GTTCTCTACTCTGGCTCGACATCTTTATTGAAGAGCAGAAATGTTATCTCTGTCTATTCATGAA	2024
Db	1722	GTTCTCTCAGTTGGCGGAATGCTCCTCACAGAGAGAACCTGATGACCGTTATCATTAAG	1781
QY	2025	ACTCTGCTAGAAAGTTTACCT-----GAGTACTTGGCAGGACAAATAAATCAACCTC	2078
Db	1782	GCTTTCATGGACCAATTTGAACACAGAGATGCCCAGGCGAGATCCAGTTTGAACGCTAC	1841
QY	2079	CAGGGTTATAGCCAGGACAAATTTGGGAAGAGTATATACGATTAATATGTGACCTAAAGTAT	2138
Db	1842	ACTGCCCTCCAAAGCCTTCAAGTTTCAGGAGAGTCCAGAGCCTCATCTTAGATCTCAAGTAT	1901
QY	2139	ATCCTGATCAGCAACCCACAATATGAGCAAGAAAGATTAAAGATGCAATTCCTTGAAGGT	2198
Db	1902	GTATTGATTAGCAAAACCAACGGAGTGTCTAGATGAGCTGAGGCAGAAAGTTCTTTACAAGGG	1961
QY	2199	TTTTCGATCTTTTGAAGATCTTACCTGTATGCGAGGGAATGGAAGAAATCCGAAGACAG	2258
Db	1962	TTCGATGCCCTTTTGGAAATTACTTGAAGTGCATGCGAGGGAATGGACCCGATCAACGCGTCAG	2021
QY	2259	GTTGGGCAACATTTGAAGTGGATCCGATTTGGGAGGCTGCCATTGCTATACAGATGCAA	2318
Db	2022	GTGGGACAGACATTTGAGATGGAGCCAGAGTGGGAAGCAGCCTTTCACACTGCGAGATGAAG	2081
QY	2319	TTGAAGAATATTTTACTCTGTTCCAAAGAGTGTGTGCTTGTGATGAAGAAACTCTTACTT	2378
Db	2082	CTGACACAGTCATCTCAATGGTGCAGGACTGGTGTGCTCTGGACGAAAGAAGTGTAAAT	2141
QY	2379	GTGGCTTATAAAGAATCTCAAAAGCTGTGATGAGGTGCATACCAAGTTTCATATCTAGT	2438
Db	2142	GAAGCTTTACAAGAAATGCTTGCTGCACACAGTGTGATCGGCGGATTTACTGATGGT	2201
QY	2439	AGCAAGACAGT---AGTACAATCGTGTGGACATAGTTTGGAAACAAAGTCTCTACAGATGA	2495
Db	2202	GAACAGCAATACACACTCAGTATTTGTGGACACTCGGTGGAAACCATCAGATACTGTGTT	2261
QY	2496	TCTCAGGATCTGTTAAGCATACATCGCCACTCTCTAGGACCTTGCCTGCTTCATGTA	2555
Db	2262	TCGCCAAGAAAAGTTAGCATCTACCTCCCAATTTCTCGCTTGTGTCAGGTTTGCATGTA	2321
QY	2556	CGTTTAAGCAGGCTGGGTGCTGTTTCAAGACTGCATGAATTTGTGTCTTTTGGAGGACTT	2615
Db	2322	TGTTAAGCAAAAGTGAAGTGGCATATAAATTTCCAGAGCTCTACTCTTAAGTGAACGTG	2381
QY	2616	CAAGTAGAGTACTAGTGGAAATATCCTTTTACGTTGTCTGGTGTGGTCCGACAGGTGTT	2675
Db	2382	AGCCCCCATGTTGATAGAACATCTCTTAGATGTCCTTGTCTTATGTGCTCAAGTGCAT	2441
QY	2676	GCTGAGATGCGCGAAGAAATGGACTGTCTCTTATTAGCCAGGTGTTTTATTACCAAGAT	2735
Db	2442	GCTGGGATGTGGAGAGAAATGGCTTCTCTCTAGTAATCAGATCTATTACTACCATAT	2501
QY	2736	GTTAAGTGCAGAGAAGAAATGTATGATAAGATATCATCATGCTTCAGATGTCGATCT	2795
Db	2502	GTGAAATGCGGGGAGAGATGTTCCGAAAGGACATAGTGATGCTTCAGACAGGTGCTCC	2561
QY	2796	TTAATGGATCCCAATTAAGTTCTTGTGTACTGGTACTTTCAGAGGTATGAACCTGCCGAGGT	2855
Db	2562	ATGATGACCCCAACCACTTCTCTGATGATCATGCTCAGCGCGTTT-----AGACAGGAT	2621
QY	2856	TTTAAACAGACCATATCTACAAA-----AGACAGGATTTGATTT	2894
Db	2622	TTCAGACAGCGCTGACTATGGGAAGAGATTCAGTTCTGAGTTTACCCATAAGGACGCTGTT	2681
QY	2895	AAACAATATATACTTAATAGAAAGAAATGCTTCAGGTCTCTCATCTATATTGTTGGGTGAG	2954
Db	2682	CACGACACACACTCTGATCGAAGAGATGCTCTACTCTATCATCATGCTTGTGGGAGAA	2741
QY	2955	CGTTATGTACTCGAGTGGGAATGTGACCAAGAGAGGTCACAAATGAGAGAAATCATTT	3014

Db	2742	AGATTCAACCTGGGGTTGGACAGTGGCTGCCACAGATGAATCAAGAGGGAGATTATC	2801
QY	3015	CACCTGCTTTGCAATGAACCCATGCCACACAGTGGCCATTGCCAAAAATTTACCTGAGAAT	3074
Db	2802	CATCAGTTGAGCATCAAGCCCTATGGCTCACAGTGAGTGGTGAAGTCTCTGCCTGAAGAT	2861
QY	3075	GAAAAATGAACACTGGCTTAGAGAACTCATAAACAAGTGGCCACATTANGAARACCA	3134
Db	2862	GAGACAGGAGACCGGCATGGAGAGCCTCATCGAGTCGGTTGACATTTCGAAGAACCT	2921
QY	3135	GGTGTATCAGCCCATGGAGTTTATGAACATAAAGATGAATCACTGAAAGACTTCAATATG	3194
Db	2922	GGCTCACAGGGCAGGATGTATGAGCTGAAGCCAGAGTGTGCCAAGAGTTCACCTG	2981
QY	3195	TACTTTTATCATTTACTCCAAACCAGCATAGCAAGGCTGAACATATGCAGAGAAGAAAGG	3254
Db	2982	TATTTTATCATTTCTCCAGGGCAGAGCAGTCCAAAGGCAGAGGAAGCTCAGCGGAAATTG	3041
QY	3255	AGAAAAACAAGAAACAAGATGAAGCATTCGGCCACCACCACTTCCTGAATTCCTGCCCT	3314
Db	3042	AAAGAGAAATAAAGAAAGATACAGCACTCCCTCCTCCGGCTTTCACCCGTTCTGCCCT	3101
QY	3315	GCTTTCAGCAAGTGATTAACCTTCTCAACTGTGATATCATCATGTACATTTCTCAGGACC	3374
Db	3102	TTGTTCCGAGTCTGGTTAAGATCTTGCACTGTGACCTCATGCTGTACATCATGGAAAGC	3161
QY	3375	GTATTTGAGGGGCAATAGACACAGATTCTTAACCTTGTCACCGAAGGGATGCTCCAAATG	3434
Db	3162	ATCCTGCAGTGGCTGTAGAGCATCACGGTCTGCCTGGTCAGAGTCCATGTACAGAGG	3221
QY	3435	GCTTTTCAVATTCGTGGCATTTGGGTTTACTAGAAGAGAACAACAGTTTCAAAGCTCCT	3494
Db	3222	GTGCTGCATTTGATCGGGATGGCTCTCCAGAAAGAGAAGACACCACTTGGAGAACGCGTG	3281
QY	3495	GAAGAAGAAGT---AACATTTGACTTTTATCATAGGCTTCAAGATTGGGAAGTTCA---	3548
Db	3282	GAAGGCACGTGCAGACCTTCACCTTCACACAGAAGATTCAAGCCTGGTAGTGACCA	3341
QY	3549	-----GCCATGAATATACAAATGCTTTTGGAAAAACTCAAAGGAATTCCTCCAGTTAGAA	3602
Db	3342	CATAACTCCCGAGCATCTTAGCTATGCTGCAGACCTTGCAGAACGCCCTCCCTGGAA	3401
QY	3603	GGCCAGAAGACATGATAAGTGGATACTTCAGATGTTTGCACACAGTGAAGCGATTAAAGA	3662
Db	3402	GOCCACAAGACATGATCAGGTGGTGTCTAAAGATGTTTAAATGCAATTAAAGAATGAAGA	3461
QY	3663	GAAAAATCTGTTTAAATTGTAGCAACCAATCAGGATCGGAATCTATTAAAGATGATGAG	3722
Db	3462	GA-----GTGTTTCATCCACGACCCCTGTGCGCGAGGCGGGGAACCAATATGGAGGAG	3515
QY	3723	ATTACTCATATAAGAAAAAGCAAGACGAAAAAGAAAGCTTGAAGCTGCTAGGCTACAT	3782
Db	3516	AGCTCAAGAGACAAGGACAAGCAGAGAGAGGAAAAAGAAAAAGCGAGATCGCCAGACTGCGC	3575
QY	3783	CGCCAGAAGATCATGGCTCAGATGTCGTGCCCTTACAGAAAAAATCTTCATTGA	3832
Db	3576	CGGGAGAAGATCATGCCGCCAGATGTTTGCAGATGCGAGCGGCACTTCATTGA	3625

RESULT 8	AI929033	682 bp	linear	EST 23-AUG-1999
LOCUS	AI929033			
DEFINITION	au54c10.y1 Schneider fetal brain 00004 Homo sapiens cDNA clone IMAGE:2519538 5' similar to TR:070481 070481 UBIQUITIN-PROTEIN LIGASE E3 COMPONENT N-RECOGNIN ;, mRNA sequence.			
ACCESSION	AI929033			
VERSION	AI929033.1	GI:5664997		
KEYWORDS	EST.			
SOURCE	Homo sapiens (human)			
ORGANISM	Homo sapiens			
	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			



	Db	301	GCATCTGTCCTCGCTGCTTTTCTGTGGGGCTATACTATTGTTCTCAGAACACATTTTGCTGCCCA	360
REFERENCE AUTHORS	Qy	5621	GGAATTTGTAAACGGGGAAGAGGTTGGAGCTTGCCATTTTTCACGCACACTTTCACCTGTGGAGC	5680
TITLE JOURNAL COMMENT	Db	361	GGAATTTGTAAACGGGGAAGAGGTTGGAGCTTGCCATTTTTCACGCACACTTTCACCTGTGGAGC	420
	Qy	5681	C GGAGTCTGCATTTTCTTA AAAATCAGAGAANTCCGAGTGGTCTCTGGTTGAAGGTAAAGC	5740
	Db	421	CGGAGTCTGCATTTTCTTA AAAAATCAGAGAAATCCGAGTGGTCTGTTGAAGGTAAAGC	480
	Qy	5741	CAGAGGCTGTGCCATATCCAGCTCCTTACTTGGATGAATATGAGAAAACACACACCCTGGSCCT	5800
	Db	481	CAGAGGCTGTGCCATATCCAGCTCCTTACTTGGATGAATATGAGAAAACACACACCCTGGSCCT	540
	Qy	5801	GAAGAGGGGCAACCCCCTTCATTTATCTCGTAGCGGTATCGGAAGCTCCATTTGGTCTG	5860
	Db	541	GAAGAGGGGCAACCCCCTTCATTTATCTCGTAGCGGTATCGGAAGCTCCATTTGGTCTG	600
	Qy	5861	GCAACAACACTGCTATTATAGAAGAGATTGCTAGAGCCAAGAGACTTAATCAGATGTTTTATT	5920
	Db	601	GCAACAACACTGCTATTATAGAAGAGATTGCTAGAGCCAAGAGACTTAATCAGATGTTTTATT	660
	Qy	5921	TGGATTCAACTGGCAGTTACTG 5942	
	Db	661	GGGATTCACCCCTGCAGNTACTG 682	
FEATURES source		RESULT 9		
		BUI55953		
		LOCUS	BUI55953	898 bp mRNA linear EST 03-SEP-2002
		DEFINITION	AGENECOURT_7966363 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:6165232	
		ACCESSION	BUI55953	5' , mRNA sequence.
		VERSION	BUI55953.1	GI:22669485
		KEYWORDS	EST.	
		SOURCE	Homo sapiens (human)	
		ORGANISM	Homo sapiens	
			Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;	
			Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
		REFERENCE	NIH-MGC http://mgc.nci.nih.gov/.	
		AUTHORS	1 (bases 1 to 898)	
		TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)	
		JOURNAL	Unpublished	
		COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgabbs-f@mail.nih.gov Tissue Procurement: ATCC/DCTD/DTF	
			'CDNA Library Preparation: Life Technologies, Inc. CDNA sequencing by: The I.M.A.G.E. Consortium (LLNL) DNA distribution by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <a href="http://image.llnl.gov">http://image.llnl.gov</a>	
			Plate: LLAM13523 row: j column: 17 High quality sequence stop: 606.	
		FEATURES	Location/Qualifiers	
		source	1 .898	
			/organism="Homo sapiens"	
			/mol_type="mRNA"	
			/db_xref="taxon:9606"	
			/clone="IMAGE:6165232"	
			/tissue_type="melanotic melanoma"	
			/lab_host="DH10B (phage-resistant)"	
			/clone_lib="NIH_MGC_72"	
			/note="Organ: skin; Vector: pCMV-SPORT6; Site: NotI; Site2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 2 kb. Library constructed by Life Technologies."	
		BASE COUNT	254 a 181 c 191 g 272 t	
		ORIGIN		
			Query Match 10.6%; Score 668.8; DB 13; Length 898;	
			Best Local Similarity 99.6%; Pred. No. 3e-127;	
			Matches 681; Conservative 0; Mismatches 2; Indels 1; Gaps 1;	

REFERENCE	1 (bases 1 to 682)
AUTHORS	Hillier,L., Allien,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S., Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
TITLE	WashU-NCI human EST Project
JOURNAL	Unpublished
COMMENT	Other_ESTS: au64c10.xl Contact: Wilson RK Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu This clone is available royalty-free through LLNL ; contact the IMAGE Consortium ( <a href="mailto:info@image.llnl.gov">info@image.llnl.gov</a> ) for further information. Possible reversed clone; similarity on wrong strand Seq primer: -4ORP from Glibco High quality sequence stop: 450.
FEATURES	Location/Qualifiers 1..682 <code>/organism="Homo sapiens"</code> <code>/mol_type="mRNA"</code> <code>/db_xref="taxon:9606"</code> <code>/clone="IMAGE:2519538"</code> <code>/sex="male"</code> <code>/tissue_type="frontal lobe"</code> <code>/lab_host="DH10B"</code> <code>/dev_stage="5 months post-conception"</code> <code>/clone_lib="Schneider fetal brain 00004"</code> <code>/notes="Organ: brain; Vector: pBluescript SK (Stratagene); Site_1: SstI; Site_2: XhoI; Double-stranded cDNA was prepared from human fetal brain tissue. 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-GAGAGAGAGAAGCTCAAGGATCCCTTAATAATTACCCCCCCCCCCC-3' and 3' adaptor sequence: 5'-GAGAGAGAGACTCGAGTGTGTTTTTTTTTTT-3'. The library was size-selected for &gt;0.5 kb inserts and has an average insert size estimated at 1.2 kb. This library was constructed using the CAP-trapper method for full-length enrichment and has not undergone amplification. Library was constructed by Dr. Claudio Schneider (LNCIB-Area Science Park, Trieste, Italy)."</code>
BASE COUNT	178 a 155 c 166 g 182 t 1 others
ORIGIN	
Query Match	10.7%; Score 674.6; DB 9; Length 682;
Best Local Similarity	99.3%; Pred. No. 1.9e-128;
Matches	677; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY	5261 AACTCCGCCCTGAGAACTGCATACCRAATTCGTCAGAGGAGGTACAGTGCACCTCTGTAG 5320
Db	1 AACTCCGCCCTGAGAACTGCATACCRAATTCGTCAGAGGAGGTACAGTGCACCTCTGTAG 60
QY	5321 CTATCTATCTTTACCTACAAAATTTGTCCTCGCTCTTCCAGGAATATGGGATACTGTAAG 5380
Db	61 CTATCTATCTTTACCTACAAAATTTGTCCTCGCTCTTCCAGGAATATGGGATACTGTAAG 120
QY	5381 GCCCTTGCTCCAGAGGTGGTGCAGATCCTCGCTTACTAACAATGTTTGAAGCAAAAAA 5440
Db	121 GCCCTTGCTCCAGAGGTGGTGCAGATCCTCGCTTACTAACAATGTTTGAAGCAAAAAA 180
QY	5441 CACGGTGGTCAGGTACCCCTAGAAAAGAAATAGTTTGATAGAGCTTCCTCATGACATATAG 5500
Db	181 CACCGTGGTCAGGTACCCCTAGAAAAGAAATAGTTTGATAGAGCTTCCTCATGACATATAG 240
QY	5501 CTCGCTCCTCAATCAAGCTTCTCATTTTCAGGTGCCACAGCTCTGCAGATCATGAGCGAAA 5560
Db	241 CTCGCTCCTCAATCAAGCTTCTCATTTTCAGGTGCCACAGCTCTGCAGATCATGAGCGAAA 300
QY	5561 GCATCCTCTGCTCTGCGCTTTTCTGTGGGGGTATACTATGTCTCTCAGAACATTTTGGCTGCCA 5620
Db	301 GCATCCTCTGCTCTGCGCTTTTCTGTGGGGGTATACTATGTCTCTCAGAACATTTTGGCTGCCA 360
QY	5621 GGAATTTGTGAACGGGGAAGAGCTTGGAGCTTGCATTTTTCACGCACACTTTCACCTGTGGAGC 5680
Db	361 GGAATTTGTGAACGGGGAAGAGCTTGGAGCTTGCATTTTTCACGCACACTTTCACCTGTGGAGC 420
QY	5681 CGGAGTCTGCATTTTCTCTTAAAAATCAGAGAAATCCGAGTGGTCTCTGGTTGAAGGTAAAGC 5740
Db	421 CGGAGTCTGCATTTTCTCTTAAAAATCAGAGAAATCCGAGTGGTCTCTGGTTGAAGGTAAAGC 480
QY	5741 CAGAGGCTGTGCCATATCCAGCTCCTTACTTGGATGAATATGGAGAAACACACACCTGGSCCT 5800
Db	481 CAGAGGCTGTGCCATATCCAGCTCCTTACTTGGATGAATATGGAGAAACACACACCTGGSCCT 540
QY	5801 GAAGAGGGGCAACCCCTTCATTTATCTCGTAGCGGTATCGGAAGCTCCATTTGGTCTG 5860
Db	541 GAAGAGGGGCAACCCCTTCATTTATCTCGTAGCGGTATCGGAAGCTCCATTTGGTCTG 600
QY	5861 GCAACAACACTGCATTTATAGAAGAGATTGCTAGAGCCAAAGAGACTAATCAGATGTTTATT 5920
Db	601 GCAACAACACTGCATTTATAGAAGAGATTGCTAGAGCCAAAGAGACTAATCAGATGTTTATT 660
QY	5921 TGGATTCAACTGGCAGTTACTG 5942
Db	661 GGGATTCACTCCAGTACTG 682
RESULT	9
LOCUS	BU155953 898 bp mRNA linear EST 03-SEP-2002
DEFINITION	AGENCOURT_7966363 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:6165232 5', mRNA sequence.
ACCESSION	BU155953
VERSION	BU155953.1 GI:22669485
KEYWORDS	EST.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
AUTHORS	NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> .
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished
JOURNAL	Contact: Robert Strausberg, Ph.D. Email: <a href="mailto:cgaabs-f@mail.nih.gov">cgaabs-f@mail.nih.gov</a> Tissue Procurement: ATCC/DCTD/DTF
COMMENT	'CDNA Library Preparation: Life Technologies, Inc. CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <a href="http://image.llnl.gov">http://image.llnl.gov</a> Plate: LLAM13523 row: j column: 17 High quality sequence stop: 606.
FEATURES	Location/Qualifiers 1..898 <code>/organism="Homo sapiens"</code> <code>/mol_type="mRNA"</code> <code>/db_xref="taxon:9606"</code> <code>/clone="IMAGE:6165232"</code> <code>/tissue_type="melanotic melanoma"</code> <code>/lab_host="DH10B (phage-resistant)"</code> <code>/clone_lib="NIH_MGC_72"</code> <code>/note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 2 kb. Library constructed by Life Technologies."</code>
BASE COUNT	254 a 181 c 191 g 272 t
ORIGIN	
Query Match	10.6%; Score 668.8; DB 13; Length 898;
Best Local Similarity	99.6%; Pred. No. 3e-127;
Matches	681; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY	5626	TTGTGAACGGGGAAGAGTTGGAGCTTGGAGCTTGCATTTTTCACGCACATTCACCTGTGGAGCCGGAG	5685
Db	12	TTGTGAACGGGGAAGAGTTGGAGCTTGCATTTTTCACGCACATTCACCTGTGGAGCCGGAG	71
QY	5686	TCGTGCATTTTCCTTAAAAATCAGAGAAATGCCGAGTGGTCCCTGGTTGAAGGTAAGGCCAGAG	5745
Db	72	TCGTGCATTTTCCTTAAAAATCAGAGAAATGCCGAGTGGTCCCTGGTTGAAGGTAAGGCCAGAG	131
QY	5746	GCTGTGCTATCCAGCTCCTTACTTGGATGAATATGAGAAACAGACCCCTGGCTGAAGA	5805
Db	132	GCTGTGCTATCCAGCTCCTTACTTGGATGAATATGAGAAACAGACCCCTGGCTGAAGA	191
QY	5806	GGGGCAACCCCTTCATTTATCTCGTGCAGCGGTATCGGAGCTCCATTTGGTCTGGCAAC	5865
Db	192	GGGGCAACCCCTTCATTTATCTCGTGCAGCGGTATCGGAGCTCCATTTGGTCTGGCAAC	251
QY	5866	AACACTGCATTTATAGAGAGATTGCTAGGAGCCAAAGACACTAATCAGATGTTATTTGGAT	5925
Db	252	AACACTGCATTTATAGAGAGATTGCTAGGAGCCAAAGACACTAATCAGATGTTATTTGGAT	311
QY	5926	TCAACTGGCAGTTACTGTAGCTCCCAACTCTGCCTCAAGACAATCACAATGACGACAGT	5985
Db	312	TCAACTGGCAGTTACTGTAGCTCCCAACTCTGCCTCAAGACAATCACAATGACGACAGT	371
QY	5986	AGTAAAGGCTGATTCAAAATATATGAAACCTTTCTGAGGCTGGGAAAGTATTGGAGGGT	6045
Db	372	AGTAAAGGCTGATTCAAAATATATGAAACCTTTCTGAGGCTGGGAAAGTATTGGAGGGT	431
QY	6046	CTTTTCTCCATGCTCCAGGTTCACTTACATCAATAAATATTTCTTAATGAGTATTGCT	6105
Db	432	CTTTTCTCCATGCTCCAGGTTCACTTACATCAATAAATATTTCTTAATGAGTATTGCT	491
QY	6106	TTCAATTAGCAACATATGCTTACAGGAAAGGACATAGATCAATCTGTTTATGTG	6165
Db	492	TTCAATTAGCAACATATGCTTACAGGAAAGGACATAGATCAATCTGTTTATGTG	551
QY	6166	CTAGTATTTCCAGGAATTTATCCCTTCATAATTTGCTCATTTTCAATTTTATTCATCC	6225
Db	552	CTAGTATTTCCAGGAATTTATCCCTTCATAATTTGCTCATTTTCAATTTTATTCATCC	611
QY	6226	ACTTGTAGATGAAGTCAAGCTCAACAGTGTAGACATTTTATGTGTT-GGTTAACTCTT	6284
Db	612	ACTTGTAGATGAAGTCAAGCTCAACAGTGTAGACATTTTATGTGTTGGGTTAACTCTT	671
QY	6285	CTGCAATTTTGTATTTGGTGTGTTT	6308
Db	672	CTGCCATTTGTATTTGGGGTTTT	695

RESULT 10  
BU618516/c  
LOCUS  
DEFINITION  
UI-H-FHL-bfk-c-21-0-UI.s1 NCI\_CGAP\_FHL Homo sapiens CDNA clone  
UI-H-FHL-bfk-c-21-0-UI 3', mRNA sequence.  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 668)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
Tissue Procurement: James Martin  
CDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa

Clone Distribution: Clone distribution information can be obtained from Dr. M. Bento Soares, Bento-soares@uiowa.edu  
Seq primer: M13 FORWARD  
POLYA=Yes.

FEATURES	Location/Qualifiers
source	1..668
	/organism="Homo sapiens"
	/mol_type="mRNA"
	/db_xref="taxon:9606"
	/clone="UI-H-FHL-bfk-c-21-0-UI"
	/tissue_type="Cell Line"
	/dev_stage="Adult"
	/lab_host="DH10B (Life Technologies)"
	/clone_lib="NCI CGAP_FHL"
	/note="Organ: Chondrosarcoma; Vector: pT73-Pac (Pharmacia I; NCI CGAP_FHL is a modified polylinker; Site_1: EcoR I; Site_2: Not a cell line derived from grade I chondrosarcoma tissue. The library was constructed and normalized according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is AGAATCCGGC. The cell line was provided by Dr. James Martin from the University of Iowa."
	TAG_LIB=UI-H-FHL
	TAG_TISSUE=Human Chondrosarcoma Cell Line CS8 - Grade 1
	Chondrosarcoma
	TAG_SEQ=AGAATCCGGC"
BASE COUNT	180 a 143 c 110 g 235 t
ORIGIN	
	Query Match 10.5%; Score 664.8; DB 13; Length 668;
	Best Local Similarity 99.7%; Pred. No. 1.9e-126;
	Matches 666; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY	2607 GAGGACTTTCAGTAGAGGTACTAGTGAATATCCTTTACGTGTCTGTGGTGGTGC 2666
Db	668 GAGGACTTTCAGTAGAGGTACTAGTGAATATCCTTTACGTGTCTGTGGTGGTGC 609
QY	2667 CAGGTTGTCTGAGATGTGGGAAGAAATGGAGTCTCTCTATTAGCCAGGTGTTTAT 2726
Db	608 CAGGTTGTCTGAGATGTGGGAAGAAATGGAGTCTCTCTATTAGCCAGGTGTTTAT 549
QY	2727 TACCAAGATGTTAAGTGCAGAGAAATGTATGATAAAGATATCATGCTTCAGATT 2786
Db	548 TACCAAGATGTTAAGTGCAGAGAAATGTATGATAAAGATATCATGCTTCAGATT 489
QY	2787 GGTGCATCTTTAATGGATCCCAATAGTCTTTGTTACTGGTACTTCAGAGGTATGAAC 2846
Db	488 GGTGCATCTTTAATGGATCCCAATAGTCTTTGTTACTGGTACTTCAGAGGTATGAAC 429
QY	2847 GCCGAGGCTTTTAAACAGACCATATCTACAAAGACCAGGATTTGATTAAACAATAAT 2906
Db	428 GCCGAGGCTTTTAAACAGACCATATCTACAAAGACCAGGATTTGATTAAACAATAAT 369
QY	2907 AACTAATAGAGAAATGCTTCAGGCTCCTCATCTATATTGTTGGGTGAGGCTTATGTACCT 2966
Db	368 AACTAATAGAGAAATGCTTCAGGCTCCTCATCTATATTGTTGGGTGAGGCTTATGTACCT 309
QY	2967 GGAGTGGGAAATGTGACCAAGAGAGGTTCACAATGAGAGAAATCATTCACCTTCCTTTGC 3026
Db	308 GGAGTGGGAAATGTGACCAAGAGAGGTTCACAATGAGAGAAATCATTCACCTTCCTTTGC 249
QY	3027 ATTGAACCCATGCCACACAGTGCCTATTCGCAAAATTTACCTGAGAAATGAAATAATGAA 3086
Db	248 ATTGAACCCATGCCACACAGTGCCTATTCGCAAAATTTACCTGAGAAATGAAATAATGAA 189

```

QY 3087 ACTGCTTAGAGATGTCATAAACAAGTGGCCACATTTAAGAAACCAGGTGTATCAGGC 3146
Db 188 ACTGCTTAGAGATGTCATAAACAAGTGGCCACATTTAAGAAACCAGGTGTATCAGGC 129
QY 3147 CATGAGTATTATGAACCTAAAGATGAATCACTGGAAGACTTCAATATGTACTTTTATCAT 3206
Db 128 CATGAGTATTATGAACCTAAAGATGAATCACTGGAAGACTTCAATATGTACTTTTATCAT 69
QY 3207 TACTCCAAACCCAGCATAGCAAGGCTGAACATATGCAGAGAGAAAAGGAGAAACAAGAA 3266
Db 68 TACTCCAAACCCAGCATAGCAAGGCTGAACATATGCAGAGAGAAAAGGAGAAACAAGAA 9
QY 3267 AACRAAGA 3274
Db 8 AACRAAAA 1

RESULT 11
BG534574 756 bp mRNA linear EST 03-APR-2001
LOCUS 602553425F1 NIH_MGC_77 Homo sapiens cDNA clone IMAGE:4663182 5',
DEFINITION mRNA sequence.
ACCESSION BG534574
VERSION BG534574.1 GI:13526116
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 756)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs@mail.nih.gov
Tissue Procurement: CLONTECH Laboratories, Inc.
cDNA Library Preparation: CLONTECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLC1465 row: a column: 07
High quality sequence stop: 751.
FEATURES
Location/Qualifiers
1..756
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4663182"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NIH_MGC_77"
/notes="organ: lung; Vector: pDNR-LIB (Clontech); Site_1:
Site_2: SfiI (ggccattagc); Site_3: SfiI (ggccattagc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CACGCCATATGGCC-3' and 3' adaptor sequence:
5'-ATTCTAGAGCGGAGCGCGGCACATG-dt(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.9
kb (range 0.5-4.0 kb). 12/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: this is a NIH_MGC Library."
BASE COUNT 274 a 138 c 152 g 192 t
ORIGIN

Query Match
Best Local Similarity 10.4%; Score 656.6; DB 10; Length 756;
Matches 730; Conservative 0; Mismatches 19; Indels 12; Gaps 5;

QY 2836 GGTATGAACCTGCCGAGCTTTTAAACAAGACCATATCTACAAAGACCAGGATTTGATTA 2895
Db 1 GGTATGAACCTGCCGAGCTTTTAAACAAGACCATATCTACAAAGACCAGGATTTGATTA 60

```

```

QY 2896 AACATATATACACTAATAGAGAATAATGCTTCAGGTCCCTCATCTATATTGTGGTGAGC 2955
Db 61 AACATATATATACACTAATAGAGAATAATGCTTCAGGTCCCTCATCTATATTGTGGTGAGC 120
QY 2956 GTTATCTACTGGAGTGGGAAATGTGACCAAAAGAGGTCAACAATCAGAGAAATCATTC 3015
Db 121 GTTATCTACTGGAGTGGGAAATGTGACCAAAAGAGGTCAACAATCAGAGAAATCATTC 180
QY 3016 ACTTGTCTTGGCATTGAACCCATGCCACACAGTGGCCATTTGCCAAAAATTTACCTGAGAA 3075
Db 181 ACTTGTCTTGGCATTGAACCCATGCCACACAGTGGCCATTTGCCAAAAATTTACCTGAGAA 240
QY 3076 AAAATAATGAACCTGGCTTAGAGAAATGTCTAATAACAAGTGGCCACATTTAAGAAACCA 3135
Db 241 AAAATAATGAACCTGGCTTAGAGAAATGTCTAATAACAAGTGGCCACATTTAAGAAACCA 300
QY 3136 GTGTATCAGGCCCATGGAGTTTATGAACCTAAAGATGAATCACTGGAAGACTTCAATATGT 3195
Db 301 GTGTATCAGGCCCATGGAGTTTATGAACCTAAAGATGAATCACTGGAAGACTTCAATATGT 360
QY 3196 ACTTTTATCATTTACTCCAAAAACCCAGCATAGCAAGGCTGAACATATGCAGAGAAAGGA 3255
Db 361 ACTTTTATCATTTACTCCAAAAACCCAGCATAGCAAGGCTGAACATATGCAGAGAAAGGA 420
QY 3256 GAAACAAGAAAACAAGATGAAG-----CATTTGGCCGCCACCACCTCCTGAAATCT 3309
Db 421 GAAACAAGAAAACAAGATGAAGTTAAAAACATTTGCCGCCACCACCTCCTGAAATCT 480
QY 3310 GCCCTGCTTTCAGCAAAAGTGAATTAACCTTCTCAACTCTGATAT-CATGATGTACATTTCT 3368
Db 481 GCCCTGCTTTCAGCAAAAGTGAATTAACCTTCTCAACTCTGATATCCATGATGTACATTTCT 540
QY 3369 AGGACCGTATTTCAGCGGGCAATAGACACAGATTCTAACTTGTGGACCGAAGGATGCTC 3428
Db 541 AGGACCGTATTTCAGCGGGCAATAGACACAGATTCTAACTTGTGGACCGAAGGATGCTC 600
QY 3429 CAAATGGCTTTTCATATCTGGCATTTGGTGGTTTACTAGAGAGAACACAGCTTCAAAAA 3488
Db 601 CAAATGGCTTTTCATATTTCTGGCATTTGGTGGTTTACTAGAGAGAACACAGCTTCAAAAA 659
QY 3489 GCTCTCGAAGAGAAGTAACATTTGACTTTTATCATTAAGCTTCAAGATTTGGGAGTTCA 3548
Db 660 GCTCTCGAAGAGAAGTAACATTTGACTTTTATCATTAAGCTTCAAGATTTGGGAGTTCA 719
QY 3549 GCCATGAATATACAAATGCTTTTGGAAAAAACAACCTCAAGGAAT 3589
Db 720 G-CATGAATATACAA---TGTTTTGGAAAAAATCAAGGATT 756

RESULT 12
AI361043/c 641 bp mRNA linear EST 15-FEB-1999
LOCUS qy03fil.x1 NCI_CGAP_Brn23 Homo sapiens cDNA clone IMAGE:2010957 3',
DEFINITION similar to TR:P91133 P91133 SIMILAR TOS. CEREBISIAE
UBIQUITIN-PROTEIN LIGASE E3 COMPONENT Sp:P19812. ; mRNA sequence.
ACCESSION AI361043
VERSION AI361043.1 GI:4112664
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 641)
AUTHORS NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute / National Institute of Neurological
Disorders and Stroke, Brain Tumor Genome Anatomy Project
(CGAP/BTGP), Tumor Gene Index
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs@mail.nih.gov
Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima

```

Bonaldo, Ph.D.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www.bio.llnl.gov/bbrp/image/image.html  
Insert length: 1083 Std Error: 0.00  
Seq primer: -400P from Gibco  
High quality sequence stop: 453.  
Location/Qualifiers

## FEATURES

source

1. .641  
/organism="Homo sapiens"  
/mol\_type="mrna"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2010957"  
/tissue\_type="glioblastoma (pooled)"  
/lab\_host="DH108"  
/clone\_lib="NCI-CGAP\_Brn23"  
/note="Organ: brain; Vector: pT73D-Pac (Pharmacia) with a  
modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st  
strand cDNA was primed with a Not I - oligo(dT) primer [5'  
TGTTACCAATCTGAAGTGGGAGCGCGCATATCTTTTCTTTTCTTTT  
T 3']; double-stranded cDNA was ligated to Eco RI  
adaptors (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of the modified pT73 vector.  
Library is normalized, and was constructed by Bento  
Soares and M.Fatima Bonaldo."

BASE COUNT 174 a 155 c 135 g 175 t 2 others

Query Match 10.1%; Score 637.4; DB 9; Length 641;  
Best Local Similarity 99.5%; Pred. No. 8.7e-121;  
Matches 638; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY	5457	CCTAGAAAAGAAATAGTTGTAGAGTTCCTGTGACTATAGTCGCTCCTGAATCAA	5516
Db	641		
QY	5517	GCTTCTCATTTCAGGTCACCGCTCTGCAGATGATGAGGAAGCATCTGCTCTGC	5576
Db	581		
QY	5577	CTTTCTGTGGGCTACTATGTTCTCAGAACATTTGCTGCCAGGAAATTTGAACGGG	5636
Db	521		
QY	5637	GAAGAGTTGGAGCTTGCAATTTTACGCACTTCTGAGCGCGGAGTCTGCATTTTC	5696
Db	461		
QY	5697	CTAAAAATCAGAAATGCCAGTGGTCTGTTGAAGGTAAGCCAGAGCTGTGCCTAT	5756
Db	401		
QY	5757	CCAGCTCCTTACTTGGATGAATATGAGAAACAGACCTCGCTGAAGAGGGCAACCCC	5816
Db	341		
QY	5817	CTTCAATTTATCTGAGCGGTATCGGAAGCTCCATTTGGTCTGGCAACACACTGCATT	5876
Db	281		
QY	5877	ATAGAAGAGATTGCTTAGGCGCCAAAGACACTAATCAGATGTTATTTGGATTCAACTGGCAG	5936
Db	221		
QY	5937	TTACTGTAGCTTCAACTCTGCTCAAGACAATCACAATAGCAGACAGTAGTAAGCTG	5996
Db	161		
QY	5997	ATTCAAAATTTATGAAATCTTTCTGAGGCTGGGAAAGTATTGGAGGGTCTTTTGTCCCA	6056
Db	101		

QY	6057	TGTCAGGTTCACTTACATCAATAAAATATTTCTTAATGGA	6097
Db	41		
RESULT 13	BU955331	774 bp	mrna
LOCUS	AGNCOURT_10609014	NIH_MGC_126	Homo sapiens cDNA clone
DEFINITION	IMAGE:6727400	5', mRNA sequence.	
ACCESSION	BU955331		
VERSION	BU955331.1	GI:24184903	
KEYWORDS	EST.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	1 (bases 1 to 774)		
TITLE	NIH-MGC http://mgc.nci.nih.gov/.		
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)		
COMMENT	Unpublished		
	Contact: Robert Strausberg, Ph.D.		
	Email: cgapbs-remail.nih.gov		
	Tissue Procurement: NCI		
	cDNA Library Preparation: Michael Brownstein Laboratory		
	DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)		
	Clone distribution: MGC clone distribution information can be		
	found through the I.M.A.G.E. Consortium/LLNL at:		
	http://image.llnl.gov		
	Plate: L10M3047 row: f column: 07		
	High quality sequence stop: 555.		
FEATURES	Location/Qualifiers		
source	1. .774		
	/organism="Homo sapiens"		
	/mol_type="mrna"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:6727400"		
	/tissue_type="mixed (pool of 40 RNAs)"		
	/lab_host="DH10B (T1-phage-resistant)"		
	/clone_lib="NIH_MGC_126"		
	/note="Vector: pDNR-LIB; Site_1: SfiI (ggccattatggcc);		
	Site_2: SfiI (ggccctcgcc); Double-stranded cDNA was		
	prepared from a pool of 40 cell line polyA+ RNAs (bladder		
	- 2%, blood - 33.4%, brain - 5.6%, breast - 12.5%, colon -		
	4%, connective tissue - 1.4%, eye - 1%, intestine - 2.6%,		
	kidney - 2.2%, liver - 5.7%, lung - 10.8%, NK-cell - 5.2%		
	, ovary - 4%, pharynx - 2.5%, prostate - 4.3%, salivary		
	gland - 1.3%, and skin - 2.3%). 5' and 3' adaptors were		
	used in cloning as follows:		
	5'-AAGCAGTGTATCAACGAGTGGCCATTACGCGGG-3' and		
	5'-ATTCTAGAGCGGCGGCGACATG-DT(30)NN-3'. Full-length		
	enriched library was constructed using the Clontech		
	Creator SMART kit and size-selected to contain the 0.5-1		
	kb size fraction (other fractions present in NIH_MGC_127		
	and NIH_MGC_128). Library created in the laboratory of T.		
	Usdin, M.D., Ph.D. (NIH, NIH). Note: this is a NIH_MGC		
	Library."		

BASE COUNT	247 a	151 c	154 g	222 t
ORIGIN				
Query Match	10.1%;	Score 635.2;	DB 13;	Length 774;
Best Local Similarity	95.2%;	Pred. No. 2.5e-120;		
Matches 678;	Conservative	0;	Mismatches 28;	Indels 6;
Gaps	2;			

QY	1375	ATTGTGCTCTTTCAATGATGAACACCATTCATATGACCGCTATACAGCTACAAA	1434
Db	3	ATTGTGCTCTTTCAATGATGAACACCATTCATATGACCGCTATACAGCTACAAA	62
QY	1435	GAGCTCTTGACTGTGAGCTCGCAGAGGCCAGTTCATACACCTGCGCATTCGCAAGAGG	1494
Db	63	GAGCTCTTGACTGTGAGCTCGCAGAGGCCAGTTCATACACCTGCGCATTCGCAAGAGG	122

```
QY 1495 GTCGTGGGCTGTAAAGCGGAGCTTATGCTGTGCGAGGAGCAAGGAAGATATAA 1554
Db 123 GTCGTGGGCTGTAAAGCGGAGCTTATGCTGTGCGAGGAGCAAGGAAGATATAA 182
QY 1555 AGAGTCATTTCAGAAAATGCTCTCAACATCCACTTCATGTAGAAGTATTACACTCAGAGA 1614
Db 183 AGAGTCATTTCAGAAAATGCTCTCAACATCCACTTCATGTAGAAGTATTACACTCAGAGA 242
QY 1615 TTATGGCTCATCAGAAATTTGCTTTGGCTCTTGGTCTGATGATGAACAAAATATFAGACT 1674
Db 243 TTATGGCTCATCAGAAATTTGCTTTGGCTCTTGGTCTGATGATGAACAAAATATFAGACT 302
QY 1675 ATTCAGTGACTTTAGGAGATCTTTTCCCAAGCATGCTTTAGAGAAACCTGACTCGG 1734
Db 303 ATTCAGTGACTTTAGGAGATCTTTTCCCAAGCATGCTTTAGAGAAACCTGACTCGG 362
QY 1735 AGAATCCCTGCTCATAGCAGGTTAATGCTTTGGATGCAAAAGCTTTTATAAGTGCC 1794
Db 363 AGAATCCCTGCTCATAGCAGGTTAATGCTTTGGATGCAAAAGCTTTTATAAGTGCC 422
QY 1795 GTAAGATCCTTCATGAATGATCTTCAGCAGTCTTTTATGGAGATGGAATACAAAAAC 1854
Db 423 GTAAGATCCTTCATGAATGATCTTCAGCAGTCTTTTATGGAGATGGAATACAAAAAC 482
QY 1855 TCTTGGCTATGGAATTTCTGAAGTATTATAACAACACTGCAGAAAGATATACAGTATG 1914
Db 483 TCTTGGCTATGGAATTTCTGAAGTATTATAACAACACTGCAGAAAGATATATACAGTATG 542
QY 1915 ATCATGACAGAAGTATCTATACTGACATCTTCAGTTCAGATGTTTACTGTTCTACTC 1974
Db 543 ATCATGACAGAAGTATCTATACTGACATCTTCAGTTCAGATGTTTACTGTTCTACTC 602
QY 1975 TGGCTCGACATCTTATGAAGCAGAAATGTTATCTCTGTCATTAAGTAACTGCTAG 2034
Db 603 TGGCTCGACATC---TTGAAAAACAGAAAGTTATCCCTGGCATTACTGGAACCTGCTC 659
QY 2035 AA---TTTACCCTGAGTACTTGGACGAGCAACATAAATTCACATTCACGG 2083
Db 660 AAAGTTTTTACCCTGAGTACTTGGACGAGCAACATAAATTCACATTCACGG 711

RESULT 14
BG862813 797 bp mRNA linear EST 29-MAY-2001
LOCUS 602799074F1 NCI_CGAP_Mam4 Mus musculus cDNA clone IMAGE:4934370 5',
DEFINITION mRNA sequence.
ACCESSION BG862813
VERSION BG862813.1 GI:14213351
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 797)
NIH-MGC http://mgc.ncl.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Priscilla Furth
Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10863 row: d column: 19
High quality sequence stop: 690.
Location/Qualifiers
1..797
/organism="Mus musculus"
```

```
/mol_type="mRNA"
/strain="NMRI"
/db_xref="taxon:10090"
/clone="IMAGE:4934370"
/issue_type="tumor, gross tissue"
/dev_stage="5 months"
/lab_host="DH10B"
/clone_lib="NCI_CGAP_Mam4"
/Note="Organ: mammary; Vector: pCMV-Sport6; Site:1: SalI;
Site:2: NotI; Cloned unidirectionally. Primer: Oligo dr.
Library constructed by Life Technologies. Investigators
providing samples: Lothar Hennighausen/Priscilla Furth,
NIH Reference for transgenic model: Li et al., Cell Growth
and Differentiation 7, 3-11 (1996)."
```

BASE COUNT 194 a 177 c 207 g 219 t  
ORIGIN

Query Match 9.9%; Score 627.2; DB 12; Length 797;  
Best Local Similarity 87.6%; Pred. No. 1.1e-118;  
Matches 697; Conservative 0; Mismatches 98; Indels 1; Gaps 1;

QY 5171 TCCTGGCTGGTATTTTGGGTCTCTCACTGAAGAAATGGCATCACCCCTTATCTTCGCTGTGC 5230  
Db 2 TCCGGGCTGGTACCTGTGCTCTCCCTGAGGAACGCATCACCCCTTACCTCCGCTGTGC 61  
QY 5231 TGCATTTGTTTCCACTATTTACTTGGGTAACCTCGCCCTGAGGAACCTGCATACCAATTC 5290  
Db 62 TGCATCTGCTTTTCCACTATTTACTTGGAGTAGTCCGCTGGAAGAACTGTTTGGCAATTC 121  
QY 5291 TGCAGAAAGAGAGTACAGTGCATCTGTAGCTATCTATCTTTACCTACAAATTTGTTCCCT 5350  
Db 122 TGCAGAGAGAGAAATTCAGTGCATCTGTAGCTATCTATCTTTACCCACAAATTTGTTCCCT 181  
QY 5351 GCTCTTCCAGGAATATTTGGGATACTGTAAAGCCCTTGTCTCCAGAGGTGGTGTGCAGATCC 5410  
Db 182 GCTTTTCCAGGAATATTTGGGATACCAATAGGCCCTTACTACAGAGTGGTGTGGAGATCC 241  
QY 5411 TGCCTTACTAACTGTTTGAAGCAAAAAACACCGTGGTGCAGTACCTCAGAAAGAAAA 5470  
Db 242 TGCCTTACTCAAGTCTTTGAAGCAGAAAAAGTGTGTGGTGCAGTACCTCAGAAAGAAAA 301  
QY 5471 TAGTTTGTAGAGCTTCTCTGATGATAGTATAGTCCCTCCTGAATCAAGCTTCTCATTTTCAG 5530  
Db 302 TAGTTTGTAGAGCTTCTCTGAGGACTACAGCTCTCTCTTAATACAGCTTCTCATTTTAG 361  
QY 5531 GTGCCACAGTCTGCAGATGATGAGCGAAAGCATCTGTCTCTGCTCTGCTTTTCTGTGGGG 5590  
Db 362 GTGTCCACGGTCTGCAGATGATGAGCGAAAGCATCTGTCTCTGCTCTGCTTCTGTGGGG 421  
QY 5591 TATACTATGTTCTCAGAACATTTGCTGCCAGGAAATTTGTCAACGGGGAAGGTTGAGC 5650  
Db 422 CATCTGTGTTCTCAGAACATCTGTTGCCAAGAAATAGTGAATGGGGAAGAGGTTGGAGC 481  
QY 5651 TTGCATTTTTCAGCAGCTTCACTGTGGCGGAGTCTGCATTTTCTCTAAATAACAGAGA 5710  
Db 482 GTGCGTTTTTCATGCGCTTCATTTGTGCTGTGAGTCTGCATTTTCTCTAAATAACAGAGA 541  
QY 5711 ATGCCAGTGGTCTGCTTGAAGTAAAGCAGAGGCTGTGCCTATCCAGCTCCTTACTT 5770  
Db 542 ATGCGGGTGGTCTGCTGTGGAAGAAAAAGCCAGAGGCTGTGCCTATCCACGCCCTTACTT 601  
QY 5771 GGATGAATATGGAAGAACAGACCCCTGGCTGAAGAGGGGCAACCCCTTCTATTATCTCG 5830  
Db 602 GGATGAATATGGAAGAACAGACCCAGGCTTAAAGAGAGGAAACCCCTTCTATTATCTCG 661  
QY 5831 TCAGCGGTATCGGAAGCTCCATTGGTC-TGGCAACAACACTGCATTATAGAGAGATTG 5889  
Db 662 GGAGCGGTATCGGAAGCTGCATTTGGTCTTGGCAAAAGGACTGGATTATAGAGAGATTG 721  
QY 5890 CTAGGAGCCCAAGAGACTAATCAGATGTTATTGGATTCAACTGSCAGTTACTGTGAGCTC 5949  
Db 722 GTCGAGCGGGAGAGACTAATCAGATGCTATTGTGACTTAAGTGGAGTTACTGTGAGATC 781



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 26, 2003, 19:40:47 : Search time 12240.1 Seconds  
(without alignments)  
17396.489 Million cell updates/sec

Title: US-09-724-126A-18  
Perfect score: 5205  
Sequence: 1 atggcgacgagagctgg.....tcaactggcagttactgtga 5205

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl:\*

- 1: gb\_ba:\*
- 2: gb\_htg:\*
- 3: gb\_in:\*
- 4: gb\_ov:\*
- 5: gb\_ov:\*
- 6: gb\_pat:\*
- 7: gb\_ph:\*
- 8: gb\_pl:\*
- 9: gb\_pr:\*
- 10: gb\_ro:\*
- 11: gb\_sts:\*
- 12: gb\_sy:\*
- 13: gb\_un:\*
- 14: gb\_vi:\*
- 15: em\_ba:\*
- 16: em\_fun:\*
- 17: em\_hum:\*
- 18: em\_in:\*
- 19: em\_mu:\*
- 20: em\_om:\*
- 21: em\_or:\*
- 22: em\_ov:\*
- 23: em\_pat:\*
- 24: em\_ph:\*
- 25: em\_pl:\*
- 26: em\_ro:\*
- 27: em\_sts:\*
- 28: em\_un:\*
- 29: em\_vi:\*
- 30: em\_htg\_hum:\*
- 31: em\_htg\_inv:\*
- 32: em\_htg\_other:\*
- 33: em\_htg\_mus:\*
- 34: em\_htg\_pln:\*
- 35: em\_htg\_rtd:\*
- 36: em\_htg\_mam:\*
- 37: em\_htg\_vrt:\*
- 38: em\_sv:\*
- 39: em\_htgo\_hum:\*
- 40: em\_htgo\_mus:\*
- 41: em\_htgo\_other:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result NO.	Score	Query Match %	Length	DB	ID	Description
1	5136	98.7	5250	9	AY061886	AY061886 Homo sapi
2	4971.6	95.5	5128	9	AF525401	AF525401 Homo sapi
3	4085.6	78.5	6395	6	AR030784	AR030784 Sequence
4	4085.6	78.5	6395	6	AR121463	AR121463 Sequence
5	4085.6	78.5	6395	10	AF061555	AF061555 Mus muscu
6	2927.4	56.2	3059	9	AX714232	AX714232 Sequence
7	2927.4	56.2	3059	9	AK056441	AK056441 Homo sapi
8	2434.4	46.8	2550	6	BD156870	BD156870 Primer fo
9	2434.4	46.8	2550	9	AK027803	AK027803 Homo sapi
10	1294.2	24.9	5268	9	AY061884	AY061884 Homo sapi
11	1151.8	22.1	5265	10	AY061885	AY061885 Mus muscu
12	996.2	19.1	1001	6	AR030785	AR030785 Sequence
13	996.2	19.1	1001	6	AR121464	AR121464 Sequence
14	994.2	19.1	999	9	AF061556	AF061556 Homo sapi
15	979.4	18.8	6158	9	AB002347	AB002347 Human mRN
16	771	14.8	3502	10	BC031403	BC031403 Mus muscu
17	650	12.5	2958	9	AK026998	AK026998 Homo sapi
18	615.4	11.8	818	9	HSX521	BD149613 Primer fo
19	534.6	10.3	712	6	BD149613	BD149613 Primer fo
20	513.4	9.9	2475	10	BC026391	BC026391 Mus muscu
21	486.2	9.3	2512	9	AK026948	AK026948 Homo sapi
22	313	6.0	1368	10	BC025617	BC025617 Mus muscu
23	258.8	5.0	190727	9	AC068724	AC068724 Homo sapi
24	246.8	4.7	807	6	BD079625	BD079625 Cancer-as
25	224	4.3	5027	10	AK122254	AK122254 Mus muscu
26	215.2	4.1	910	10	MM1UBR2	AF067372 Mus muscu
27	215.2	4.1	168200	2	AC120631	AC120631 Rattus no
28	215.2	4.1	198946	10	AL844548	AL844548 Mouse DNA
29	215.2	4.1	235978	2	AC094209	AC094209 Rattus no
30	212.2	4.1	107304	2	AC016274	AC016274 Homo sapi
31	212.2	4.1	164468	2	AC021899	AC021899 Homo sapi
32	212.2	4.1	166518	9	AC090514	AC090514 Homo sapi
33	195	3.7	6381	3	AY094815	AY094815 Drosophil
34	191.6	3.7	800	6	BD079626	BD079626 Cancer-as
35	180.8	3.5	164468	2	AC021899	AC021899 Homo sapi
36	160.6	3.1	3980	9	HS1UBR4	AF067383 Homo sapi
37	160.2	3.1	3411	9	AK090874	AK090874 Homo sapi
38	159	3.1	455	6	BD071823	BD071823 Secreted
39	157.6	3.0	594	11	G79301	G79301 S210P604SRG
40	150.2	2.9	107304	2	AC016274	AC016274 Homo sapi
41	146.2	2.8	181602	10	AL935168	AL935168 Mouse DNA
42	132	2.5	28834	2	AC017982	AC017982 Drosophil
43	132	2.5	152545	3	AC010922	AC010922 Drosophil
44	132	2.5	179016	3	AC018489	AC018489 Drosophil
45	132	2.5	298641	3	AE003504	AE003504 Drosophil

ALIGNMENTS

RESULT 1  
AY061886  
LOCUS AY061886 5250 bp mRNA linear PRI 30-DEC-2002  
DEFINITION Homo sapiens ubiquitin ligase E3 alpha-I mRNA, complete cds.  
ACCESSION AY061886  
VERSION AY061886.1 GI:27434483  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
AUTHORS 1 (bases 1 to 5250)  
TITLE Han, H. Q. and Kwak, K.  
FULL-length human ubiquitin ligase E3 alpha-I (E3 alpha)  
JOURNAL Unpublished





|||||  
1381 TTCCAGGTTATAGCCAGGACAAATGGGAAGATATATGCGATATAATGTGACCTAAAG 1440  
QY  
1396 TATATCTGATCAGCAAAACCCAAATATATGACAGAAAGATTAAAGATGAGTTCCTTTGAA 1455  
Db  
1441 TATATCTGATCAGCAAAACCCAAATATATGACAGAAAGATTAAAGATGAGTTCCTTTGAA 1500  
QY  
1456 GGTTCGATCTTTTGAAGATTCCTACCTGTATGACAGGGAATGGAAGAAATCCGAAGA 1515  
Db  
1501 GGTTCGATCTTTTGAAGATTCCTACCTGTATGACAGGGAATGGAAGAAATCCGAAGA 1560  
QY  
1516 CAGGTGGCAACACATGGAAGTGGATCCTGATGGGAGGCTGCCATTCCTATACAGATG 1575  
Db  
1561 CAGGTGGCAACACATGGAAGTGGATCCTGATGGGAGGCTGCCATTCCTATACAGATG 1620  
QY  
1576 CAATTGAAGAAATATTTACTCATGTGCCAAGAGTGGTGTGTCATCAAGAACTCTTA 1635  
Db  
1621 CAATTGAAGAAATATTTACTCATGTGCCAAGAGTGGTGTGTCATCAAGAACTCTTA 1680  
QY  
1636 CTTGTGGCTTATAAGAAATGTCACAAAGCTGTGATGAGTGCAGTACCAAGTTTCATATCT 1695  
Db  
1681 CTTGTGGCTTATAAGAAATGTCACAAAGCTGTGATGAGTGCAGTACCAAGTTTCATATCT 1740  
QY  
1696 AGTAGACAGACAGTAGTACAAATCGTGTGGACATAGTTTGGAAACAAAGTCCCTACAGATG 1755  
Db  
1741 AGTAGACAGACAGTAGTACAAATCGTGTGGACATAGTTTGGAAACAAAGTCCCTACAGATG 1800  
QY  
1756 TCTGAGGATCTTGTAAAGATACATCTGCCACTCTAGGACCCCTGCTGGTCTTCATGTA 1815  
Db  
1801 TCTGAGGATCTTGTAAAGATACATCTGCCACTCTAGGACCCCTGCTGGTCTTCATGTA 1860  
QY  
1816 CGTTTAAAGCAGGCTGGTGTCTTTCAAGACTGTCATGAATTTGTCTTTTTCAGGACTTT 1875  
Db  
1861 CGTTTAAAGCAGGCTGGTGTCTTTCAAGACTGTCATGAATTTGTCTTTTTCAGGACTTT 1920  
QY  
1876 CAAGTAGAGGTACTAGTGGAAATATCCTTTACGTTCTCTGGTGTGGTGGCCAGGTTGTT 1935  
Db  
1921 CAAGTAGAGGTACTAGTGGAAATATCCTTTACGTTCTCTGGTGTGGTGGCCAGGTTGTT 1980  
QY  
1936 GCTGAGATGTGCCGAAGAAATGGACTGTCTCTTATTTAGCCAGGTGTTTATTTACCAAGAT 1995  
Db  
1981 GCTGAGATGTGCCGAAGAAATGGACTGTCTCTTATTTAGCCAGGTGTTTATTTACCAAGAT 2040  
QY  
1996 GTTAAGTCAGAGAAGAAATGTATGATAAAGATATCATCTGCTTCAGATTTGGTGCATCT 2055  
Db  
2041 GTTAAGTCAGAGAAGAAATGTATGATAAAGATATCATCTGCTTCAGATTTGGTGCATCT 2100  
QY  
2056 TTAATGGATCCCAATAAGTCTTGTACTGGTACTTTCAGAGGTATGAACCTTGCCGAGGCT 2115  
Db  
2101 TTAATGGATCCCAATAAGTCTTGTACTGGTACTTTCAGAGGTATGAACCTTGCCGAGGCT 2160  
QY  
2116 TTTAAACAGACATATCTACAAAGACCAAGGATTTGATTAACAAATATAATACACTAATA 2175  
Db  
2161 TTTAAACAGACATATCTACAAAGACCAAGGATTTGATTAACAAATATAATACACTAATA 2220  
QY  
2176 GAAGAAATGCTTCAGGCTCTCATCTATATTTGGGTGAGCGTTATGTACCTGGAGTGGGA 2235  
Db  
2221 GAAGAAATGCTTCAGGCTCTCATCTATATTTGGGTGAGCGTTATGTACCTGGAGTGGGA 2280  
QY  
2236 AATGTGACCAAGAGAGGTGCACATGAGAGAAATCATCTGCTTTTGCATTTGAAGCC 2295  
Db  
2281 AATGTGACCAAGAGAGGTGCACATGAGAGAAATCATCTGCTTTTGCATTTGAAGCC 2340  
QY  
2296 ATGCCACAGTGCCTATGCCCCAAATTTTACCTGAGATGAAAATTAATGAAACTTGGCTTA 2355  
Db  
2341 ATGCCACAGTGCCTATGCCCCAAATTTTACCTGAGATGAAAATTAATGAAACTTGGCTTA 2400  
QY  
2356 GAGAAATGTCATAAACAAGTGGCCACATTTAAGAAACAGAGTGTATCAGGCCATGGAGTT 2415  
Db  
2401 GAGAAATGTCATAAACAAGTGGCCACATTTAAGAAACAGAGTGTATCAGGCCATGGAGTT 2460  
QY  
2416 TATGAACCTAAAAGATGATCACTCAAGACTTCAATATGTACTTTTATCATTTACTCCAAA 2475  
|||||

Db  
2461 TATGAACCTAAAAGATGAATCACATGGAAGACTTCAATATATGTACTTTTATCATTTACTCCAAA 2520  
QY  
2476 ACCAGCATAGCAAGGCTGAACATATGCGAAGAAAAGGAGAAAACAAGAAAACAAGAT 2535  
Db  
2521 ACCAGCATAGCAAGGCTGAACATATGCGAAGAAAAGGAGAAAACAAGAAAACAAGAT 2580  
QY  
2536 GAAGCAATGGCGCCACACCACTCTGAAATTCCTGACCTTGCCTTGTTCAGCAAAAGTGATTAAC 2595  
Db  
2581 GAAGCAATGGCGCCACACCACTCTGAAATTCCTGACCTTGCCTTGTTCAGCAAAAGTGATTAAC 2640  
QY  
2596 CTTCTCAACTGTGATATCATGTATGATTCCTCAGGACCGTATTTTCAGCGGCAATAGAC 2655  
Db  
2641 CTTCTCAACTGTGATATCATGTATGATTCCTCAGGACCGTATTTTCAGCGGCAATAGAC 2700  
QY  
2656 ACAGATTTCAACTTTGTGGACCGAAGGATGCTCCAAATGGCTTTTTCATATTTCTGGCATTTG 2715  
Db  
2701 ACAGATTTCAACTTTGTGGACCGAAGGATGCTCCAAATGGCTTTTTCATATTTCTGGCATTTG 2760  
QY  
2716 GGTTTACTAGAAGAGAAGCAACAGCTTCAAAAAGCTCTCTGAAGAAGAAAGTAACATTTGAC 2775  
Db  
2761 GGTTTACTAGAAGAGAAGCAACAGCTTCAAAAAGCTCTCTGAAGAAGAAAGTAACATTTGAC 2820  
QY  
2776 TTTTATCATAGAGCTTCAAGATTTGGAAAGTTTGGAAAGTTTGGAAAGTTTGGAA 2835  
Db  
2821 TTTTATCATAGAGCTTCAAGATTTGGAAAGTTTGGAAAGTTTGGAAAGTTTGGAA 2880  
QY  
2836 AAACCTCAAAGGAATTCCTCCAGTTAGAAGCCAGAGGACATGATAACGTGGATACTTCAG 2895  
Db  
2881 AAACCTCAAAGGAATTCCTCCAGTTAGAAGCCAGAGGACATGATAACGTGGATACTTCAG 2940  
QY  
2896 ATGTTTGCACACAGTGAAGCGATTAAAGAAAAATCTTTTAAATTTAGTACACACCATCA 2955  
Db  
2941 ATGTTTGCACACAGTGAAGCGATTAAAGAAAAATCTTCTTAAATTTAGTACACACCATCA 3000  
QY  
2956 GGATCGGAATCTATTAAAGAAATGATGAGATTAATCATGATATAAAGAAAAGCAGAACGAAA 3015  
Db  
3001 GGATCGGAATCTATTAAAGAAATGATGAGATTAATCATGATATAAAGAAAAGCAGAACGAAA 3060  
QY  
3016 AGAAAAGCTGAAGCTGTAGGCTACATGCCAGAGATCATGGCTCAGATGTCTGCCTTA 3075  
Db  
3061 AGAAAAGCTGAAGCTGTAGGCTACATGCCAGAGATCATGGCTCAGATGTCTGCCTTA 3120  
QY  
3076 CAGAAAACCTTCATTGAAACTCATAACTCATGTATGACAAATACATCAGAAATGCTGGG 3135  
Db  
3121 CAGAAAACCTTCATTGAAACTCATAACTCATGTATGACAAATACATCAGAAATGCTGGG 3180  
QY  
3136 AAAGAAGATTCOATTTATGGAGGAAGAGACACCCAGCAGTCAGTGACTTCTAGAATT 3195  
Db  
3181 AAAGAAGATTCOATTTATGGAGGAAGAGACACCCAGCAGTCAGTGACTTCTAGAATT 3240  
QY  
3196 GCTTTGGGTCTTAAACGGGGTCCATCTGTTACTGAAAAGGAGGTGCTGACGTGCATCCCT 3255  
Db  
3241 GCTTTGGGTCTTAAACGGGGTCCATCTGTTACTGAAAAGGAGGTGCTGACGTGCATCCCT 3300  
QY  
3256 TGCCAAGAGAACAGSAGSTGAAAATAGAAAATAATGCCATGTAATTCGSCCTGTGTC 3315  
Db  
3301 TGCCAAGAGAACAGSAGSTGAAAATAGAAAATAATGCCATGTAATTCGSCCTGTGTC 3360  
QY  
3316 CAGAAATCTACTGCCCTTAACCCAGCACAGGGGAAAACCCATAGAACTCTCAGGAGAAAGCC 3375  
Db  
3361 CAGAAATCTACTGCCCTTAACCCAGCACAGGGGAAAACCCATAGAACTCTCAGGAGAAAGCC 3420  
QY  
3376 CTAGACCCACTTTTTCATGGATCCAGACTTGGCATATGGAACCTTATACAGGAAGCTGTGGT 3435  
Db  
3421 CTAGACCCACTTTTTCATGGATCCAGACTTGGCATATGGAACCTTATACAGGAAGCTGTGGT 3480  
QY  
3436 CATGTAATGCAGCAGTGTGCTGGCAGAAAGTATTTTGAAGCTGTACAGCTGAGCTCTCAG 3495  
Db  
3481 CATGTAATGCAGCAGTGTGCTGGCAGAAAGTATTTTGAAGCTGTACAGCTGAGCTCTCAG 3540  
QY  
3496 CAGGCATTCATGTTGACCTTTTTCACCTTGGAAAGTGGAGAAATATCTTTGCCCCTCTTTC 3555  
Db  
3541 CAGGCATTCATGTTGACCTTTTTCACCTTGGAAAGTGGAGAAATATCTTTGCCCCTCTTTC 3600

QY 3556 AAATCTCTGTGCAATACTGTGATCCCAATATTTCCTTTGCAACCTCAAAAGATAAAGCAGT 3615  
|||||  
Db 3601 AAATCTCTGTGCAATACTGTGATCCCAATATTTCCTTTGCAACCTCAAAAGATAAAGCAGT 3660  
|||||  
QY 3616 GAGAATGCAGATGCTCTGTCTCAACTTTTGACCCCTGGCAGGGTGATACAGACTGTCTCTG 3675  
|||||  
Db 3661 GAGAAATGCAGATGCTCTGTCTCAACTTTTGACCCCTGGCAGGGTGATACAGACTGTCTCTG 3720  
|||||  
QY 3676 GCCAGAAATACAGGTATATAATAAGACATGCTAAAGAGAGAAACCCAAATTCCTATTTC 3735  
|||||  
Db 3721 GCCAGAAATACAGGTATATAATAAGACATGCTAAAGAGAGAAACCCAAATTCCTATTTC 3780  
|||||  
QY 3736 TTTTAATCAAGAAATGGAGATCTTACTTTGGAGTTCCATTCCTCAGTTTGGGGTT 3795  
|||||  
Db 3781 TTTTAATCAAGAAATGGAGATCTTACTTTGGAGTTCCATTCCTCAGTTTGGGGTT 3840  
|||||  
QY 3796 GAGTCTTCGATTAATAATTCAAATAGCATCAAGGAAATGGTTATTCCTTTGCCCACAACA 3855  
|||||  
Db 3841 GAGTCTTCGATTAATAATTCAAATAGCATCAAGGAAATGGTTATTCCTTTGCCCACAACA 3900  
|||||  
QY 3856 ATTTATAGAATTTGGAATGGAAGTCCACCTGATGAAAGGGATCTCGAGTCCCAATGCTG 3915  
|||||  
Db 3901 ATTTATAGAATTTGGAATGGAAGTCCACCTGATGAAAGGGATCTCGAGTCCCAATGCTG 3960  
|||||  
QY 3916 ACCTGGAGCACCTGGCTTTTCACTATCCAGGCAATTTGAAATCTATTGGGAGATGAAGGA 3975  
|||||  
Db 3961 ACCTGGAGCACCTGGCTTTTCACTATCCAGGCAATTTGAAATCTATTGGGAGATGAAGGA 4020  
|||||  
QY 3976 AAACCTCTGTTTGAGGACACTTCAAAATAGGAGCATATGTCCTGAAAGCATTAATGCG 4035  
|||||  
Db 4021 AAACCTCTGTTTGAGGACACTTCAAAATAGGAGCATATGTCCTGAAAGCATTAATGCG 4080  
|||||  
QY 4036 TTTTSCAGTTGCACAGAGATTACCTGCTCCTCAGTCTCTGATACAGAAACATCTGTTCTG 4095  
|||||  
Db 4081 TTTTSCAGTTGCACAGAGATTACCTGCTCCTCAGTCTCTGATACAGAAACATCTGTTCTG 4140  
|||||  
QY 4096 CTTCATATCAGTTGTTCTTCAACATAAAATCAGAAGATACACATGCTCTGCTATATA 4155  
|||||  
Db 4141 CTTCATATCAGTTGTTCTTCAACATAAAATCAGAAGATACACATGCTCTGCTATATA 4200  
|||||  
QY 4156 GATCTGTTTCATGTTTGGTGGGTGCTGTGTAGCATTCCTCATTCCTTATTTGGGATGAC 4215  
|||||  
Db 4201 GATCTGTTTCATGTTTGGTGGGTGCTGTGTAGCATTCCTCATTCCTTATTTGGGATGAC 4260  
|||||  
QY 4216 CCTGTTGATCTGCAGCCTTCTTCAGTTAGTTTCTTCTATACCAACCTTTATCTCTCCAT 4275  
|||||  
Db 4261 CCTGTTGATCTGCAGCCTTCTTCAGTTAGTTTCTTCTATACCAACCTTTATCTCTCCAT 4320  
|||||  
QY 4276 TTGATCACCATGGCACACATGCTTTCAGATCTACTTACAGTAGACAGGCTACCCCTT 4335  
|||||  
Db 4321 TTGATCACCATGGCACACATGCTTTCAGATCTACTTACAGTAGACAGGCTACCCCTT 4380  
|||||  
QY 4336 GCTCAGGTTCAAGAAGACAGTGAAGAGCTCATTCCTCGCATCTCTTTCTTTCGCAAAAT 4395  
|||||  
Db 4381 GCTCAGGTTCAAGAAGACAGTGAAGAGCTCATTCCTCGCATCTCTTTCTTTCGCAAAAT 4440  
|||||  
QY 4396 TCTCAATATACAAGTGGCTCCATTTGGGTGTGATATTCCTGGCTGTGATTTTGGGTCTCA 4455  
|||||  
Db 4441 TCTCAATATACAAGTGGCTCCATTTGGGTGTGATATTCCTGGCTGTGATTTTGGGTCTCA 4500  
|||||  
QY 4456 CTGAAGAAATGGCATACCCCTTATCTTGGCTGTGCTGCATTTTTCACATATTTACTT 4515  
|||||  
Db 4501 CTGAAGAAATGGCATACCCCTTATCTTGGCTGTGCTGCATTTTTCACATATTTACTT 4560  
|||||  
QY 4516 GGGTAACTCCGCTGAGGAAGTGCATACCAATTCGAGAGGAGATACAGTCCACTC 4575  
|||||  
Db 4561 GGGTAACTCCGCTGAGGAAGTGCATACCAATTCGAGAGGAGATACAGTCCACTC 4620  
|||||  
QY 4576 TGTAGCTATCTATCTTTACCTACAAATTTGTTCTGCTCTTCCAGGAATATTTGGGATACT 4635  
|||||  
Db 4621 TGTAGCTATCTATCTTTACCTACAAATTTGTTCTGCTCTTCCAGGAATATTTGGGATACT 4680  
|||||

QY 4636 GTAAGGCCCTTGTCTCCAGAGCGGTGTGCAGATCCTGCTTACTAACTGTTTGAAGCAA 4695  
|||||  
Db 4681 GTAAGGCCCTTGTCTCCAGAGGTGTGTGCAGATCCTGCTTACTAACTGTTTGAAGCAA 4740  
|||||  
QY 4696 AAAAACACCGTGGTGCAGTACCCCTAGAAAAAGAAATAGTTTGATAGAGTTCCTCATGAC 4755  
|||||  
Db 4741 AAAAACACCGTGGTGCAGTACCCCTAGAAAAAGAAATAGTTTGATAGAGTTCCTCATGAC 4800  
|||||  
QY 4756 TATAGTCCCTCTGTAATCAAGCTTCTCATTTTCAGTGGCCACGCTGTGCAGATGATGAG 4815  
|||||  
Db 4801 TATAGTCCCTCTGTAATCAAGCTTCTCATTTTCAGTGGCCACGCTGTGCAGATGATGAG 4860  
|||||  
QY 4816 CGAAAGCATCTCTGCTCTGCTCTTCTGTGGGCTATATACTATGTTTCTCAGAACATTTGC 4875  
|||||  
Db 4861 CGAAAGCATCTCTGCTCTGCTCTTCTGTGGGCTATATACTATGTTTCTCAGAACATTTGC 4920  
|||||  
QY 4876 TGCCAGGAAATTTGTAACGGGGAAGAGTTTGAGGCTTGCAATTTTTCAGCACTTCACTGT 4935  
|||||  
Db 4921 TGCCAGGAAATTTGTAACGGGGAAGAGTTTGAGGCTTGCAATTTTTCAGCACTTCACTGT 4980  
|||||  
QY 4936 GGAGCCGAGTCTGCATTTTCTTAAATTCAGAAATGCCGAGTGGTCTGCTTGAAGGT 4995  
|||||  
Db 4981 GGAGCCGAGTCTGCATTTTCTTAAATTCAGAAATGCCGAGTGGTCTGCTTGAAGGT 5040  
|||||  
QY 4996 AAAGCCAGAGGCTGTCCCTATCCAGCTCCTTACTTGGATGAATATGGAGAAACACACCT 5055  
|||||  
Db 5041 AAAGCCAGAGGCTGTCCCTATCCAGCTCCTTACTTGGATGAATATGGAGAAACACACCT 5100  
|||||  
QY 5056 GGCCTGAAGAGGGGCAACCCCTTCATTTATCTCGTAGCGGTATCGGAAGCTCCATTTG 5115  
|||||  
Db 5101 GGCCTGAAGAGGGGCAACCCCTTCATTTATCTCGTAGCGGTATCGGAAGCTCCATTTG 5160  
|||||  
QY 5116 GTCTGGCAACACACTGCATTTATAGAAGATTTGTAGGAGCCAGAGACTAATCAGATG 5175  
|||||  
Db 5161 GTCTGGCAACACACTGCATTTATAGAAGATTTGTAGGAGCCAGAGACTAATCAGATG 5220  
|||||  
QY 5176 TTATTTGGATTCAACTGGCAGTTACTGTGA 5205  
|||||  
Db 5221 TTATTTGGATTCAACTGGCAGTTACTGTGA 5250  
|||||

RESULT 2  
AF525401  
LOCUS AF525401 5128 bp mRNA linear PRI 01-JAN-2003  
DEFINITION Homo sapiens UBR1 E3a ligase mRNA, partial cds.  
ACCESSION AF525401  
VERSION AF525401.1 GI:27451603  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Cranial; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 5128)  
AUTHORS Dgany, O., Avidan, N., Delaunay, J., Krasnov, T., Shalmon, L.,  
Shalev, H., Eidelitz-Markus, T., Kapelushnik, J., Cattian, D.,  
Pariente, A., Tulliez, M., Cretien, A., Schischmanoff, P.O.,  
Iolascon, A., Fibach, E., Koren, A., Rossler, J., Le Merrier, M.,  
Yaniv, I., Zalizov, R., Ben-Asher, E., Olender, T., Lancet, D.,  
Beckmann, J.S. and Tamary, H.  
TITLE Congenital dyserythropoietic anemia type I is caused by mutations  
in codanin-1  
JOURNAL Am. J. Hum. Genet. 71 (6), 1467-1474 (2002)  
MEDLINE 22340442  
PUBMED 12434312  
REFERENCE 2 (bases 1 to 5128)  
AUTHORS Ben-Asher, E.  
TITLE Direct Submission  
JOURNAL Submitted (27-JUN-2002) Molecular Genetics, The Weizmann Institute  
of Science and the Schneider Medical Center, Hertzl Street,  
Rehovot 76100, Israel  
FEATURES  
source 1. 5128  
/organism="Homo sapiens"

```
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="15"
/map="15q13"
/cell_type="erythroid"
<1..>5128
/codon_start=1
/product="UBR1 E3a ligase"
/protein_id="AA014997.1"
/db_xref="GI:27451604"
/translation="LQPTPORLASWDDQVDFYAFHLHLAQLVPELIFAEIMDPLEK
QESVMSITPLEWLFEGDPDLCLEKHSFAQLCGRVFKSGEYITSCRDCAIDP
TCLVMDQFDSVQHNHRYKMTSTGGFCDCGDEAMTKPFCVNPHEPGRAGIKEN
SRCLNEEIVQARKIFFPSVYKYVEMTIIWEERKELPELQIREKNEYCVLFDNFH
HSYDHYISLORALDELAQAOLHTAIDEGRAVKAGAVAAQOEAKEDIKSHENV
SOHLAVEHLSIMAHOKFALRLFSWMNKIMSYSSDFRQIFCOACI.REBPSDENPCL
ISRLMLADKLYKGARKHELIFSSFFMEMEYKKLFAMEFVYIKQLOKEYISDDHD
RSISLAVQMTPTPLAHLHLEBNVISVITETLEVLPEYLDLRNKNFQGSOD
KLGRVYAVICDLKYLISKPTIWRERLMQFEGFRSLKILTCMQMEETIRQVGH
IEYDPWEAAIAQOMLKNILLMFQEWACDELLIYAYKECHKAVRCSSTFSSSK
TVQSGHSLSTKSYRVSDELVSIIHLPLSRLTAGLHVRLSGLGAVSRHREVSPEDEQ
VEVLVEYPLRLVLAQVVAEMRRNGLSLSQVFFYQDYKCREMYDKDIIMLQIGA
SLMDPNKFLLLVORLELAENFKTISTKQDILQYNTLIEMLQVLLIIVGERYP
GVGNVKEETVREIIHLICIEPHPSATAKNPENNETGLENIKNVATPEKPGV
SGHGYVELDESILDNMFYHYSKTQHSKAEMQKRRKQENKDEALPPPPPEPCP
AFSKVINLLNCDIMTYILRTVFERAIDTDSNLTWEGMOWAFHILALGLEEKQOLQ
APEEVTDFYKASRLGSSAMNIMLEKLGIPOLEGQKMTITWLOMDFVRLR
EKSLIVATSSSEKINDEITHDEKAEKRAEAARLHRQIKMAQSLQKNFIET
HKLMMNPAKQKSDSIMEESTPAVSDYSRIALGPKRPSYTEREVLVILCEQEQ
EYKINNAWLSACVQSTALTOHRGKPIELSGEALDPLFMDPLDLYGTGTSGCHVM
HAWCQKFEAVLSQORIHVDFLESGEYLCPLCKSLCNTVPIPILOPKINSE
NADALAOLITLARWITVLARISYNIHAKGENPIPIFNQMGDSTLEHFSILSG
VSSIKYSIKEMVILPATIYRIGLVKPPDERDPVPMLTWSTCAFTQAENLGL
DSGKPLFGLNQHRHGLKALMOWAQRITCPQVLIQKHLVRLLSVLLNKSEDTTP
CULSIDFLVNLVLAFLPSLWDDPVDLPQSVSSYNNHLYLFLHITMAHMLQILIT
VDTPLQAVQESDEBAHSASFAEISQYTSIGSDIGDIPGWLVWSLKNIGITPYLR
AALFFRYLLGVTPPELHTNSAREYSALCSYLSLXNLLFLOEYDWTVRPLIOLRC
ADPALLNCKOKNTVYVPRKNSLIELPDDYSCILLNOASHFRCPRAEDDERKHPVLC
LFGGAILCSNCCQIEVINGEVGACIFHAFHAGAGVCIPLKRECRVILVEGRKRC
ATPAPLDEIYGEFDDPKRGNPLJHFIYFTEAPLVRQHC1"
BASE COUNT 1540 a 1010 c 1137 g 1439 t 2 others
```

CDS

Query Match 95.5%; Score 4971.6; DB 9; Length 5128;  
 Best Local Similarity 98.6%; Pred. No. 0;  
 Matches 5060; Conservative 11; Mismatches 12; Indels 48; Gaps 4;

Db	361	TACAAGATGCATACCTTCTACTGGAGAGGGTTCTGTGACTGTGGAGACACAGAGGATGG	420
Qy	472	AAAACTGCCCTTTTGTGTAATCATGAACCTGGAAGACGAGGTACTATAAAGAGAAT	531
Db	421	AAAACTGCCCTTTTGTGTAATCATGAACCTGGAAGACGAGGTACTATAAAGAGAAT	480
Qy	532	TCACGCTGTCGGTGAATGAAGAGGTAATGTCCAAGCAGGAGAAATATTTCTTCAGTG	591
Db	481	TCACGCTGTCGGTGAATGAAGAGGTAATGTCCAAGCAGGAGAAATATTTCTTCAGTG	540
Qy	592	ATAAATATGTCGTAAGAAATGACTATATGGGAAGAGGAGAAAGAACTGCTCTGACATC	651
Db	541	ATAAATATGTCGTAAGAAATGACTATATGGGAAGAGGAGAAAGAACTGCTCTGACATC	600
Qy	652	CAGATAAGGKNR-----	666
Db	601	CAGATAAGGAGAGAAATGAAAGATACATTTGTGTCTCTTCAATGATGAACACCATCCA	660
Qy	667	VNDHSHYDHTCATATACAGCCTACAAAGAGCTCTTGTACTGTGAGCTCGCAGAGGCCAG	726
Db	661	TATGACCACGTCATATACAGCCTACAAAGAGCTCTTGTACTGTGAGCTCGCAGAGGCCAG	720
Qy	727	TGTCATACCACTGCCATTGACAAAGAGGTCGTCGGGCTGTAAAGCGGAGGCTTATGCT	786
Db	721	TTCATACCACTGCCATTGACAAAGAGGTCGTCGGGCTGTAAAGCGGAGGCTTATGCT	780
Qy	787	GCTTCCAGGAGCAAGGAGATATAAAGATCATTAAGAGTCTTCAAGAAATGTCTCTCAACATCA	846
Db	781	GCTTCCAGGAGCAAGGAGATATAAAGATCATTAAGAGTCTTCAAGAAATGTCTCTCAACATCA	840
Qy	847	CTTCATGTAGAGTATTACACATCAGAGATTATGGCTCATCAGAAATTTCTCTTGGCTCTT	906
Db	841	CTTCATGTAGAGTATTACACATCAGAGATTATGGCTCATCAGAAATTTCTCTTGGCTCTT	900
Qy	907	GTTTCTGGATGAACAAATTTATGAGCTATTCAAGTGTACTGTAGGAGAGTCTTTTGCCAA	966
Db	901	GTTTCTGGATGAACAAATTTATGAGCTATTCAAGTGTACTGTAGGAGAGTCTTTTGCCAA	960
Qy	967	GCATGCTTAGAGAGAACCTGACGTCGGAGATCCCTGTCTCATAGCAGGTTAATGCTT	1026
Db	961	GCATGCTTAGAGAGAACCTGACGTCGGAGATCCCTGTCTCATAGCAGGTTAATGCTT	1020
Qy	1027	TGGGATCAAAAGCTTTATAAAGTCCCGTAGATCCTTCATGAATGTATCTTCAGCAGT	1086
Db	1021	TGGGATCAAAAGCTTTATAAAGTCCCGTAGATCCTTCATGAATGTATCTTCAGCAGT	1080
Qy	1087	TTTTTTATGGAGATGGAATACAAAACCTCTTTGTCTATGGAATTTGTGAAGTATTATAA	1146
Db	1081	TTTTTTATGGAGATGGAATACAAAACCTCTTTGTCTATGGAATTTGTGAAGTATTATAA	1140
Qy	1147	CAACTGCAGAAAGATATATCAGTGATGATCATGACAGAGTATCTCTATAACTGCACAT	1206
Db	1141	CAACTGCAGAAAGATATATCAGTGATGATCATGACAGAGTATCTCTATAACTGCACAT	1200
Qy	1207	TCAGTTGATGTTTACTGTTCTCTACTCTGGCTCGACATCTTATTTGAAGAGCAGATGTT	1266
Db	1201	TCAGTTGATGTTTACTGTTCTCTACTCTGGCTCGACATCTTATTTGAAGAGCAGATGTT	1260
Qy	1267	ATCTCTGTCATTACTGAAACTCTGCTAGAAGTTTTACCTGAGTACTTTGGACAGGACAAT	1326
Db	1261	ATCTCTGTCATTACTGAAACTCTGCTAGAAGTTTTACCTGAGTACTTTGGACAGGACAAT	1320
Qy	1327	AAATTCAACTCCAGGGTTATAGCCAGGACAAATTTGGGAAGAGTATATGACGTAATGAT	1386
Db	1321	AAATTCAACTCCAGGGTTATAGCCAGGACAAATTTGGGAAGAGTATATGACGTAATGAT	1380
Qy	1387	GACCTAAAGTATATCTCTGATCAGCAAAACCCACAAATATGACAGAGAAATTAAGATGCA	1446
Db	1381	GACCTAAAGTATATCTCTGATCAGCAAAACCCACAAATATGACAGAGAAATTAAGATGCA	1440
Qy	1447	TTCTTTGAAGGTTTTCGATCTTTTGAAGATCTTCTTACCTGTATGAGGAAATGGAAGA	1506
Db	1441	TTCTTTGAAGGTTTTCGATCTTTTGAAGATCTTCTTACCTGTATGAGGAAATGGAAGA	1500

Qy	52	TTACCCAGACCCCTCAGGCTCTGGCATCTTGGTGGGATCAGCAAGTTGATTTTACT	111
Db	1	TTACCCAGACCCCTCAGGCTCTGGCATCTTGGTGGGATCAGCAAGTTGATTTTACT	60
Qy	112	GCTTTCTTCATCATTTGGCACAATTTGGTCCAGAAATTTACTTTGCTGAATGGACCA	171
Db	61	GCTTTCTTCATCATTTGGCACAATTTGGTCCAGAAATTTACTTTGCTGAATGGACCA	120
Qy	172	GACTTGGAAAACGAGGAGAAAGTGTACAAATGTCATATTTCACTCCACTGGAATGGTAC	231
Db	121	GACTTGGAAAACGAGGAGAAAGTGTACAAATGTCATATTTCACTCCACTGGAATGGTAC	180
Qy	232	TTATTTGGAGAACATCCAGATATTTGCTTAGAGAAATTTGAAGACACATGGAGCATTTTCA	291
Db	181	TTATTTGGAGAACATCCAGATATTTGCTTAGAGAAATTTGAAGACACATGGAGCATTTTCA	240
Qy	292	CTTTGTTGGAGGGTTTCAAAAGTGGAGAGACAACTTATTTCTTGCAGGATTTGCAATT	351
Db	241	CTTTGTTGGAGGGTTTCAAAAGTGGAGAGACAACTTATTTCTTGCAGGATTTGCAATT	300
Qy	352	GATCCACATGTGCTATGATGAGCTGCTCCAGACAGTGTTCATATAAATCATCGT	411
Db	301	GATCCACATGTGCTATGATGAGCTGCTCCAGACAGTGTTCATATAAATCATCGT	360
Qy	412	TACAAGATGCATCTTCTACTGGAGGAGGTTCTGTGACGTGTGGAGACACAGAGGATGG	471

QY	2587	GTGATTAACCTTCTCAACTGTGTGATATCATGATGTACATTTCTCAGGACCGCTATTGTGACGG	264
DB	2581	GTGATTAACCTTCTCAACTGTGTATATCATGTATCATTTCTCAGGACCGCTATTGTGACGG	2640
QY	2647	GCAATAGACACAGATTTCTAACTTGTGGACCGAAGGGATGCTCCAAATGGCTTTTCATATT	2706
DB	2641	GCAATAGACACAGATTTCTAACTTGTGGACCGAAGGGATGCTCCAAATGGCTTTTCATATT	2700
QY	2707	CTGCGATTGGTTTTACTAGAAGAGAACAACAGCTTCCAAAAGCTCCTGAAGAAGAAGTA	2766
DB	2701	CTGCGATTGGTTTTACTAGAAGAGAACAACAGCTTCCAAAAGCTCCTGAAGAAGAAGTA	2760
QY	2767	ACATTTGACTTTTAATCAATAGGCTTCAAGATTTGGGAAGTTTCAGCCATGAATATACAAATG	2826
DB	2761	ACATTTGACTTTTAATCAATAGGCTTCAAGATTTGGGAAGTTTCAGCCATGAATATACAAATG	2820
QY	2827	CTTTTGGAAAACHTCAAGGAATTCGCCAGTTTGAAGCCAGAGGACATGATAACGTGG	2886
DB	2821	CTTTTGGAAAACHTCAAGGAATTCGCCAGTTTGAAGCCAGAGGACATGATAACGTGG	2880
QY	2887	ATACTTCAGATGTTTGCACAGTGAAGCGATTAAAGAGAAAAATCTGTTTAAATGTGACGA	2946
DB	2881	ATACTTCAGATGTTTGCACAGTGAAGCGATTAAAGAGAAAAATCTGTTTAAATGTGACGA	2940
QY	2947	ACCACATCAGGATCGGAATCTATTAAAGATGATGAGATTACTCATGATATAAGAAAAAGCA	3006
DB	2941	ACCACATCAGGATCGGAATCTATTAAAGATGATGAGATTACTCATGATATAAGAAAAAGCA	3000
QY	3007	GAACGAAAAAGAAAGCTGAAGCTGTAGGCTACATCGCCAGAGATCATGGCTCAGATG	3066
DB	3001	GAACGAAAAAGAAAGCTGAAGCTGTAGGCTACATCGCCAGAGATCATGGCTCAGATG	3060
QY	3067	TCTGCCTTACAGAAAAAATCTTCAATTGAAACTCATAACTCATGTATGACAAATACATCAGAA	3126
DB	3061	TCTGCCTTACAGAAAAAATCTTCAATTGAAACTCATAACTCATGTATGACAAATACATCAGAA	3120
QY	3127	ATGCTTGGGAAAGAAGATTCCATTATGGAGAGAGAGACCCCGACAGTCAGTGACTAC	3186
DB	3121	ATGCTTGGGAAAGAAGATTCCATTATGGAGAGAGAGACCCCGACAGTCAGTGACTAC	3180
QY	3187	TCTAGAATTGCTTTGGTGCTTAAACGGGTCCATCTGTTACTGAAAAGGAGGTGCTCAGC	3246
DB	3181	TCTAGAATTGCTTTGGTGCTTAAACGGGTCCATCTGTTACTGAAAAGGAGGTGCTCAGC	3240
QY	3247	TGCAATCTTTGCCAAGAAGAACAGGAGGTGAAAAATAGAAAATATGCCATGGTATTATCG	3306
DB	3241	TGCAATCTTTGCCAAGAAGAACAGGAGGTGAAAAATAGAAAATATGCCATGGTATTATCG	3300
QY	3307	GCCTGTGCCAAAATCTACTGCTTAAACCCAGCAGAGGGAACCCATAGAACTCTCA	3366
DB	3301	GCCTGTGCCAAAATCTACTGCTTAAACCCAGCAGAGGGAACCCATAGAACTCTCA	3360
QY	3367	GGAGAAGCCTTAGACCACATTTTCATGGATCCAGACTTGGCATATGAACTTATACAGGA	3426
DB	3361	GGAGAAGCCTTAGACCACATTTTCATGGATCCAGACTTGGCATATGAACTTATACAGGA	3420
QY	3427	AGCTGTGGTCATGTAATGCACGAGTGTGCTGGCAGAAATTTTGAAGCTGTACAGCTG	3486
DB	3421	AGCTGTGGTCATGTAATGCACGAGTGTGCTGGCAGAAATTTTGAAGCTGTACAGCTG	3480
QY	3487	AGCTCTCAGCAGCGCATTCATGTTGACCTTTTTCGATTTGAAAAGTGAGAAATATCTTTGC	3546
DB	3481	AGCTCTCAGCAGCGCATTCATGTTGACCTTTTTCGATTTGAAAAGTGAGAAATATCTTTGC	3540
QY	3547	CCTCTTTGCAAAATCTCTGTGCAATCTGATCCCAATTAATTCCTTTGCAACCTCAAAAG	3606
DB	3541	CCTCTTTGCAAAATCTCTGTGCAATCTGATCCCAATTAATTCCTTTGCAACCTCAAAAG	3600
QY	3607	ATAAACAGTGAGAAATGCAGATGCTTCTTGCTCAACTTTTGACCTTGGCACGGTGGATACAG	3666
DB	3601	ATAAACAGTGAGAAATGCAGATGCTTCTTGCTCAACTTTTGACCTTGGCACGGTGGATACAG	3660
QY	3667	ACTGTTCTGGCCGAATATCAGGTTTAAATATAAGACATGCTAAAGAGAGAAAAACCAATT	3726

D	b		4741	CCTGATGACATATAGTGCCTCCTGGAAATCAAGCTTCATTTCAGGTGCCACCGGTCTGCA	4800
Q	y		4807	GATGATCAGCGAAGACATCTCTGCTGCTCTGCTGCTCTTTCTGTGGGGCTATACTATGTTCTCAG	4866
D	b		4801	GATGATCAGCGAAGACATCTCTGCTGCTCTTTCTGTGGGGCTATACTATGTTCTCAG	4860
Q	y		4867	AACAATTTGCTGCCAGGAANAATGTGAACGGGGAGAAGGTTGGAGCTTGCCATTTTTCAGCCA	4926
D	b		4861	AACAATTTGCTGCCAGGAANAATGTGAACGGGGAGAAGGTTGGAGCTTGCCATTTTTCAGCCA	4920
Q	y		4927	CTTCACATGTGGAGCGCGAGTCTGCATTTTCTTAAAAATCAGAGAATGCCAGTGGTCTCTG	4986
D	b		4921	CTTCACATGTGGAGCGCGAGTCTGCATTTTCTTAAAAATCAGAGAATGCCAGTGGTCTCTG	4980
Q	y		4987	GTTGAAGGTAAGCCACAGAGGCTGTGCTATCCAGCTCTTTACTTTGGATGAATATGGAGAA	5046
D	b		4981	GTTGAAGGTAAGCCACAGAGGCTGTGCTATCCAGCTCTTTACTTTGGATGAATATGGAGAA	5040
Q	y		5047	ACAGACCCTGGCCTGAAGAGGGGCAACCCCTTCATTTATCTCGTAGCGGTATCGGAAG	5106
D	b		5041	ACAGACCCTGGCCTGAAGAGGGGCAACCCCTTCATTT-TCTCGTAGCGGTA-CGGAG	5098
Q	y		5107	CTCCATTTGCTGTGGCAACAACACTGCAATTA	5137
D	b		5099	CTCCA- TTGGTNGGCAACAACACTGCAATTA	5128
RESULT 3					
AR030784					
LOCUS AR030784 6395 bp DNA linear PAT 29-SEP-1999					
DEFINITION Sequence 1 from patent US 5861312.					
ACCESSION AR030784					
VERSION AR030784.1 GI:5943998					
KEYWORDS					
SOURCE Unknown.					
ORGANISM Unclassified.					
REFERENCE 1 (bases 1 to 6395)					
AUTHORS Varshavsky,A. and Kwon,Y.Tae.					
TITLE Nucleic acid encoding mammalian UBRI					
JOURNAL Patent: US 5861312-A 1 19-JAN-1999;					
FEATURES					
source Location/Qualifiers					
1..6395					
BASE COUNT 1802 a 1376 c 1519 g 1698 t					
ORIGIN					
Query Match 78.5%; Score 4085.6; DB 6; Length 6395;					
Best Local Similarity 86.6%; Pred. No. 0;					
Matches 4567; Conservative 9; Mismatches 629; Indels 69; Gaps 4;					
Q	y	1	ATGCGCACGAGGAGGCTGGAGGTACTGAGGATGAAAATCAGCGCGGAGTTACCCCAG	60	
D	b	115	ATGCGCAGCAGAGATGGAGCGGCCGACAGAGATGACGCCGAGACCTCCCTG	174	
Q	y	61	ACCCTCAGGCTCTGGCATCTTGGTGGGATCAGCAAGTTGATTTTTATACTGCTTTCTTG	120	
D	b	175	GCCCCGACGCGCGGCATCGTGGTGGATCAGCAAGTTGATTTCTATCTGCTTTCTTA	234	
Q	y	121	CATCATTTGGCACAATTGGTCCAGAAATTTACTTTCCTGAATGGACCCAGACTTGGAA	180	
D	b	235	CATCATTTGGCACAATTAGTGGCAGAAATTTATTTGCTGAGATGGACCCAGATTTGGAA	294	
Q	y	181	AACGAGGAGNAAGTGACAAATCTCAATATTCACCTCACTGGAATCGTACTTATTGGA	240	
D	b	295	AAGCAAGAGAGAGTGACAGATGTCAATACTCACTCTTTGGAGTGGTACTTATTGGA	354	
Q	y	241	GAAGATCCAGATATTTCCTTAGAGAAATTTGAAGCACAGTGGAGCATTTACGCTTTGTGG	300	
D	b	355	GAGATCCGATATTTCCTTAGAGAAATTTAAACACASTGGAGGTTCCAGTTCTGTGGG	414	
Q	y	301	AGGTTTTTAAAAGTGGAGACAACTATTCTTTGACGGGATTTGTGCAATTTGATCCAACA	360	



Db 415 AAGGTTTCAAAAGTGGAGAAACAACATATTCCTGTAGGATTGTGCAATTGATCCACAG 474  
Qy 361 TGTGTACTCTGTATGGACTGTTCCAGACAGAGTGTTCATATAAAATCATCGTTACAGATG 420  
Db 475 TGTGTCTCTGTATGGACTGTTCCAAAGTAGTGTTCATAAAAACCATCGTTACAAGATG 534  
Qy 421 CATACTTCTACTGGAGGGTCTGTGACTGTGGACACAGAGGATGGAACACTGGC 480  
Db 535 CATACTTCTACTGGAGGGGCTTGTGACTGTGGACACAGAGCGTGGAAAACACTGGC 594  
Qy 481 CCTTTTGTGTAATGAACCTGGAAAGCAGGTACTATAAAGAGAAATTCACGCTGT 540  
Db 595 CCTTTTGTGTAATGAACCTGGAAAGCAGGTACTATAAAGAGAAATTCACGCTGT 654  
Qy 541 CCGTTGAATGAAGAGTAAATGTCCAAAGCCAGGAAATATTCCTTCAGTGTATAAATAT 600  
Db 655 CCATTGAATGAAGAGTAAATGTCCAAAGCCAGGAAATATTCCTTCGTTGATATAAATAC 714  
Qy 601 GTCGTAGAATGACTATATGGGAAGAGGAAAGAAAGTGCCTCTGTAAGTCCAGATAAG 660  
Db 715 ATTGTAGAATGACTATATGGGAAGAGGAAAGAAATTCCTCTGAACTGCAGATAAG 774  
Qy 661 KNR-----YVCVNDHHSYDH 675  
Db 775 GAGAAAATGAACGATACTATTGTGCTCTTTTCAACGATGAGCACCATTCGTATGATCAT 834  
Qy 676 GTCATATACAGCCTACAAGAGCTCTTGACGTGTGAGCTCGCAGAGGCCAGTTCATAC 735  
Db 835 GTGATCTACAGCTCTGCAGAGAGCTTAGATTGCGAGCTTCGAGAGGCACAGCTGCACAG 894  
Qy 736 ACTGCCATTCACAAGAGGGTCTCGGGCTGTTAAAGCGGAGCTTATGCTGCTTTCGAG 795  
Db 895 ACTGCCATTCACAAGAGGGTCTCGGGCTGTTCAAAAGCAGGTGTGTATGCCACTTGGCAG 954  
Qy 796 GAACCAAGGAAGATATAAAGAGTCATCAGAAATGCTCTCAACATCCACTTCATCTA 855  
Db 955 GAAGCAAGGAGGATATAAAGAGTCATCAGAAACGCTCTCAGCAGCCCCCTCATGTG 1014  
Qy 856 GAAGTATTACACTCAGAGATTATGGCTCATCAGAAATTTGCTTCGCTCTTGGTTCCTGG 915  
Db 1015 GAAGTGTGCACTCGTGGTTATGGCTCACCAGAAATTCGCTCTCGGCTTGGCTCCTGG 1074  
Qy 916 ATGAACAAATTTATGAGCTATTCAGTGAAGTCTTTAGGCAGATCTTTTGGCCAGCATGCCT 975  
Db 1075 ATGAACAAATTTATGAGCTATTCAGTGAAGTCTTTAGCAGATATTTTGGCAGGCTGCCTC 1134  
Qy 976 AGAGAAAGCTGACTCGGAGAAATCCCTGTCTCATAGCAGGTAACTGCTTGGGATGCA 1035  
Db 1135 GTAGAAGAACTGGCTCTGAAATCCCTGTCTTATAGCAGACTAACTGCTTGGGATGCA 1194  
Qy 1036 AAGCTTTATAAGGTGCCGTAAAGATCCTCATGAATTTGATCTTCAGCAGTTTTTTTATG 1095  
Db 1195 AAACCTTTATAAGGTGCCGTAAAGATCCTCATGAATTTGATCTTTAGTAGTTTTTTATG 1254  
Qy 1096 GAGATGGAATACAAAAACTCTTGTCTATGGAATTTGTGAAGTATTTATAAACAACACTGCAG 1155  
Db 1255 GAGATGGAATACAAAAACTCTTGTCTATGGAATTTGTGAAGTATTTATAAACAACACTGCAG 1314  
Qy 1156 AAAGAATATACAGTATGATCATGACAGAGTATCTCTATTAACCTGCACCTTTCAGTTTCAG 1215  
Db 1315 AAAGAGTACATCAGCGAGGACCCAGAGAGAGCATCTCCATTAACCGCCCTGTCCTGGCAG 1374  
Qy 1216 ATGTTTACTGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1275  
Db 1375 ATGCTCACCGTCCCGACCTTGCGCCGGCATCTTATTAAGAGCAGAAATGTTATTTCTGTC 1434  
Qy 1276 ATTACTGAACTCTCTAGAGTCTTACCTGAGTACTTGGCAGGAGAAATATAAATTCAC 1335  
Db 1435 ATTACTGAACTCTCTAGAGTCTTACCTGAGTACTTGGCAGGAGAAATATAAATTCAC 1494  
Qy 1336 TTCCAGGTTATAGCCAGGACAAATTTGGAGAGAGTATATGAGTAAATGTGACCTAAG 1395  
Db 1495 TTCCAGGTTATAGCCAGGACAAATTTGGAGAGAGTATATGAGTAAATGTGACCTAAG 1554

Qy 1396 TATATCCTGATCAGCAAAACCCACATATATGACAGAAAGATTAAAGATGCAGTTCCTTGAA 1455  
Db 1555 TATATCCTGATAGCAAGCCTGTCTATATGACAGAAAGATTAAAGATGCAGTTCCTTGAA 1614  
Qy 1456 GGTTCCTGATCTTTTGAAGATCTTACCTGTATGACGGGATGAAGAAATTCGCAAGA 1515  
Db 1615 GGTTCCTGATCTTTTGAAGATCTTACCTGTATGACGGGATGAAGAAATTCGCAAGA 1674  
Qy 1516 CAGGTGGGCAACACATTGAAGTGTCTGATTTGGAGGCTGCCATTCGTATACAGATG 1575  
Db 1675 CAAGTTGGACACACATTGAAGTGTCTGATTTGGAGGCTGCCATTCGTATACAGATG 1734  
Qy 1576 CAATTTGAAGATATTTTACTCATGTTCCAAAGTGGTGTCTGATGAAGAACTCTTA 1635  
Db 1735 CAACATAAAGATATTTTGTCTCATGTTCCAAAGTGGTGTCTGATGAAGATCTCTTA 1794  
Qy 1636 CTTGTGGCTTATAAAGATGTCAAAAGCTGTGATGAGTGCAGTACCAGTTTCATATCT 1695  
Db 1795 CTGTTGGCTTATAAAGATGTCAAAAGCTGTGATGAGTGCAGTACCAGTTTCATATCT 1854  
Qy 1696 AGTAGCAAGACAGTAGTAGTACAATCGTGTGACATAGTTTGGAAACAAGCTCTACAGAT 1755  
Db 1855 AGTACCAAGACAGTAGTAGTACAATCGTGTGACATAGTTTGGAAACAAGCTCTACAGAT 1914  
Qy 1756 TCTGAGGATCTTGTAAAGCATACATTCGCACCTCTCTAGACCCCTGCTGGTCTTCATGTA 1815  
Db 1915 TCTGAGGACCTTGTAAAGCATACATTCGCACCTCTCTAGACCCCTGCTGGTCTTCATGTA 1974  
Qy 1816 CGTTTAAAGCAGCTGGTGTCTTTCAAGCTGCATGAATTTGTCTTTTGAGGACTTT 1875  
Db 1975 CGTTTAAAGCAGCTAGTGTCTTTCAAGCTGCATGAATTTGTCTTTTGAGGACTTT 2034  
Qy 1876 CAAGTAGGTTACTAGTGAATATCTTTACGTTGTCTGTTGGTGGTCCAGGTTGTT 1935  
Db 2035 CAAGTAGGTTCTGTTGGAGTACCGCTGCGCTGCTGGTCTGGTGGCTCAGTTGTT 2094  
Qy 1936 GCTGAGATGTGCGGAAGAAATGGACTGTCTCTTATTTAGCCAGGTGTTTATTTACCAAGAT 1995  
Db 2095 GCTGAGATGTGCGGAAGAAATGGACTGTCTCTTATTTAGCCAGGTGTTTATTTACCAAGAT 2154  
Qy 1996 GTTAAGTGCAGAGAGAAATGTATGATAAGATATCATGCTTCAGATTGTTGATCT 2055  
Db 2155 GTTAAGTGCAGAGAGAAATGTATGATAAGATATCATGCTTCAGATTGTTGATCT 2214  
Qy 2056 TTAATGGATCCCAATTAAGTTCTTGTACTGGTACTTCAGAGTATGAACCTTGCAGGCT 2115  
Db 2215 ATAATGGATCCCAATTAAGTTCTTGTACTGGTACTTCAGAGATATGAACCTTACTGATCT 2274  
Qy 2116 TTTAACAAGACCATATCTACAAAGACCCAGGATTTGATTAAACAATATATACACTAATA 2175  
Db 2275 TTTAACAAGACCATATCTACAAAGACCCAGGATTTGATTAAACAATATATACACTAATA 2334  
Qy 2176 GAAGAAATGCTTCAGGCTCTCATCTATATTGGGTGAGGTTATGTACCTGGAGTGGGA 2235  
Db 2335 GAAGAAATGCTTCAGGCTCTCATCTATATTGGGGAAGCGTTATGTACCTGGAGTGGGA 2394  
Qy 2236 AATGTGACCAAGAGAGGCTCAATGAGAGAAATCATTCACCTTGTCTTCATTGAACCC 2295  
Db 2395 AATGTGACCAAGAGAGGCTCAATGAGAGAAATCATTCACCTTGTCTTCATTGAACCC 2454  
Qy 2296 ATGCCACAGTGCCTATTCGCAAAATTTTACCTGAGAATGAAATATGAACTGGCTTA 2355  
Db 2455 ATGCCACAGTGCCTATTCGCAAAATTTTACCTGAGAATGAAATATGAACTGGCTTA 2514  
Qy 2356 GAGATGTCTAATAACAAAGTGGCCACATTTAAGAAACCCAGGTGTATCAGGCCATGGAGTT 2415  
Db 2515 GAGATGTCTAATAACAAAGTGGCCACATTTAAGAAACCCAGGTGTATCAGGCCATGGAGTT 2574  
Qy 2416 TATGAATTAAGATGAATCACTGAAAGACTTCAATATGACTTTTATCTACTCTCAAA 2475  
Db 2575 TATGAATTAAGATGAATCACTGAAAGACTTCAATATGACTTTTATCTACTCTCAAA 2634



[illegible]

Db 4795 CTTTTCCAGGAATATTGGGATACCATAGGCCCTTACTACAGAGGTGGTGTGGAGATCCT 4854

QY 4672 GCCTTACTAACTGTTTGAAGCAAAAACACCGTGGTCTAGGTACCCGTAGAAAAGAAAT 4731

Db 4855 GCCTTACTCAAGTCTTTGAAGCAGAAAAGTGTGTGGTCAAGTACCCGTAGAAAAGAAAT 4914

QY 4732 AGTTTGATAGAGTCTCTGATGACTATAGTGTGCTCCTGTAATCAAGTCTTCTCATTTCCAGG 4791

Db 4915 AGTTTGATAGAGTCTCTGAGGACTACAGTGTCTTCTTAATCAGGCTTCTCACTTTAGG 4974

QY 4792 TGCCCAAGGTCACAGATGATAGGCAAAAGCACTCTGCTCCTGCTCTTTCTGTGGGGCT 4851

Db 4975 TGCCCAAGGTCACAGATGATAGGCAAAAGCACTCTGCTCCTGCTCTTTCTGTGGGGCT 5034

QY 4852 ATACTATGTTCTCAGAACATTTGCTGCCAGGAAATGTGAACGGGGAAGAGTGTGGAGCT 4911

Db 5035 ATCCTGTGTCTCAGAACATCTGTGCCAAGAAATAGTGAATGGGAAGAGTGTGGAGCG 5094

QY 4912 TGCATTTTTCACGCACTTCACTGTGGAGCCGGAGTGTGCATTTTCTTAAATCAGAGAA 4971

Db 5095 TGGCGTTTTCATCGGCTTCAATTTGTGTGGAGTGTGCATTTTCTTAAATCAGAGAA 5154

QY 4972 TGCCGAGTGTCTGTTGAAGGTAAAGCCAGAGGCTGTGCTATCCAGCTCCTTACTTGT 5031

Db 5155 TGCAAGGTGTCTGTTGGGAAGAAAGCCAGAGGCTGTGCTACCCAGCCCTTACTTGT 5214

QY 5032 GATGAATATGGAGAAACAGACCTTGGCTTGAAGGGGCAACCCCTTCAATTTATCTCGT 5091

Db 5215 GATGAATATGGAGAAACAGACCCAGGCTTAAAGAGAGAAACCCACTTCAATTTATCTCGG 5274

QY 5092 GAGCGGTATCGGAAGTCCATTTGGTGTGGCAACACACTGCAATATAGAGAGATTTGCT 5151

Db 5275 GAGCGGTATCGGAAGTCCATTTGGTGTGGCAACAGCACTGCAATATAGAGAGATTTGCT 5334

QY 5152 AGGAGCAACAGACTAATCAGATGTTATTGGATTCAACTGGCAGTTACTGTGA 5205

Db 5335 CGGAGCCAGAGACTAATCAGATGCTATTGGATTCAACTGGCAGTTACTGTGA 5388

RESULT 4

AR121463

LOCUS AR121463 6395 bp DNA linear PAT 16-MAY-2001

DEFINITION Sequence 1 from patent US 6159732.

ACCESSION AR121463

VERSION AR121463.1 GI:14105039

KEYWORDS

SOURCE Unknown.

ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 6395)

AUTHORS Varshavsky,A. and Kwon,Y.Tae.

TITLE Nucleic acid encoding mammalian Ubr1

JOURNAL Patent: US 6159732-A 1 12-DEC-2000;

FEATURES

source

1..6395

/organism="unknown"

BASE COUNT 1802 a 1376 c 1519 g 1698 t

ORIGIN

Query Match 78.5%; Score 4085.6; DB 6; Length 6395;

Best Local Similarity 86.6%; Pred. No. 0;

Matches 4567; Conservative 9; Mismatches 629; Indels 69; Gaps 4;

QY 1 ATGGCGACAGGAGGCTGGAGGTACTCAGAGGATGGAATCAGCGGAGCTTACCCCG 60

Db 115 ATGGCGACAGGAGATGGAGCGCGGAGAGGATGGACGTGAGCCCGGAGCTCCCGCTG 174

QY 61 ACCCTCAGGCTCTGGCATCTTGGGATCAGCAAGTTGATTTTATCTGCTTCTTG 120

Db 175 GCCCGCAGCGCGGATCGTGGTGGATCAGCAAGTTGATTTTATCTGCTTCTTA 234

QY 121 CATCATTTGGCACAATTTGGTCCAGAAATTTTACTTTGCTGAAATGGACCCAGACTTGGAA 180

Db 235 CATCATTTGGCACAATTAGTGCCAGAAAATTTATTTTGTGAGATGGACCCAGATTTGGAA 294

QY 181 AAGCAGGAGGAAGTGTACAAATGTCAATATTCACTCCACTGGAATGGTACTTATTGGGA 240

Db 295 AAGCAAGAAGAGGTGTACAGATGTCAATACTCACTCCTTTGGAGTGGTACTTATTGGGA 354

QY 241 GAAGATCCAGATATTGCTTTAGAGAAATTTGAAGCACAGTGGAGCAATTTTCAGCTTTGTGGG 300

Db 355 GAGGATCCGGATATTGCTTTAGAGAAATTTAAACACAGTGGAGCGTTCCAGTGTGTGGG 414

QY 301 AGGGTTTTCAAAAGTGGAGAGACAACTTATCTTCAGGGATTTGTGCAATTTGATCCAACA 360

Db 415 AAGTTTTCAAAAGTGGAGAAACAACATATTCTCTAGGATTTGTGCAATTTGATCCAAGG 474

QY 361 TGTGTACTCTCTATGGACTGCTTCCAGGACAGTGTTCATAAAATCATCGTTTACAAGATG 420

Db 475 TGTGTCTCTCTATGGACTGCTTCCAAAGTAGTGTTCATAAAACCATCGTTTACAAGATG 534

QY 421 CATACTTCTACTGGAGGAGGTTCTGTGACTGTGGAGACACAGAGCATGGAAAACCTGGC 480

Db 535 CATACTTCTACTGGAGGAGGTTCTGTGACTGTGGAGACACAGAAAGCGGTGAAAACCTGGC 594

QY 481 CCTTTTGTGTAAATCATGAACCTGGAAGAGCAGGTACTATATAAGAGAGAAATTCACGCTGT 540

Db 595 CCTTTTGTGTGATCAGAGCGCTGGAAGAGCAGGTACTACAAAGAGAGCTTACATTCG 654

QY 541 CCGTTGAATGAAGAGGTAAATTTCCAAAGCCAGGAAATATTTTCTTCACTGATATAAATAT 600

Db 655 CCATTTGAATGAAGAGGTGATTGCTCAAGCCAGGAGAAATATTCCTTCGCTGATATAAATAC 714

QY 601 GTCGTAGAATGACTATATGGGAAGAGGAAAGAACTGCCTTCTTGAACCTCAGATTAAGG 660

Db 715 ATTCTAGAATGACTATATGGGAAGAGGAAAGAAATTCGCTTCTTGAACCTCAGATTAAGG 774

QY 661 KNR-----YVCVNDHSHYDH 675

Db 775 GAGAAAATGAACGATACTATTGTGCTCTTTTCAAGGATGAGCACCATTCTGATGATCAT 834

QY 676 GTCATATACAGCCTACAAAGAGCTCTTGACTGTGAGCTGCGAGAGGCCAGTTGTCATACC 735

Db 835 GTGATCTACAGTCTCAGAGAGCTCTAGATTGCGAGCTTGCAGAGGACACAGCTGCACAG 894

QY 736 ACTGCCATTGACAAAGAGGCTCGTGGGCTGTTAAAGCGGGAGCTTATGCTGCTGCCAG 795

Db 895 ACTGCCATCGACAAAGAGGCTCGCGGCTGTCAAAGCAGAGTGTGATGCCACTTCCAGG 954

QY 796 GAAGCAAAAGAGATATAAGAGTCAATTCAGAAAATGTCTCTCAACATCCACTTCATGTA 855

Db 955 GAAGCAAAAGAGGATATAAGAGTCACTCAGAGAACGTCTCTCAGCACCCTCCATGTG 1014

QY 856 GAAATATTACACTCAGAGATTTATGGCTCATCAGAAAATTTGCTTTGCGTCTTGGTTCCTGG 915

Db 1015 GAAATGCTGCATCCGCTGTTATGGCTCACCAGAAATTCGCTCTGCGCTTGGCTCCTGG 1074

QY 916 ATGAACAAAATTTATGAGCTATTCAAGTACTTTAGGCAGATCTTTTGGCCAAAGCATGCCIT 975

Db 1075 ATGAACAAAATTTATGAGCTATTCAAGTACTTTAGACAGATATTTTGGCCAGGCTCCCTC 1134

QY 976 AGAGAAAGACCTGACTCGGAGAAATCCCTCTCTCATAGAGGTTAATGCTTTGGGATGCA 1035

Db 1135 GTAGAAGAACCTGGGCTCTGAAAATCCCTGCTTTAAGCAGAGCAATGCTTTGGGATGCA 1194

QY 1036 AAGCTTTATAAAGGTGCCCGTAAGATCCCTTCATGAATTTGATCTTCAGCAGTTTTTTTATG 1095

Db 1195 AAACHTTATAAAGGTGCCCGTAAGATCCCTTCATGAATTTGATCTTTAGTAGTTTTTATG 1254

QY 1096 GAGATGGAATACAAAACACTCTTTGCTATGGAATTTGTGAAGTATTATATAAACACTGCAG 1155

Db 1255 GAGATGGAATACAAAACACTCTTTGCTATGGAATTTGTGAAGTATTATATAAACACTGCAG 1314

QY 1156 AAAGAATATATCAGTGTGATCATGACAGAAAGTATCTCTATATACTGCATTTCACTTTCAGTTCAG 1215

Db 1315 AAAGAGTACATCAGCAGACGACCAGAGAGAGCATCTCCATAACCGCCCTGTCCTGGTGCAG 1374

QY	1216	ATGTTTACTGTCTTACTCTGCTCGACATCTTATTGAAGAGCAGAAATGTTATCTCTGTC	1275	2296	ATGCCACACAGTGGCCATTGGCCAAAAATTTACCTGAGAAATGAAAAATTAATGAACTGGCTTA	2355
Db	1375	ATGCTCACCGTCCCGACCTTGGCCGGCACTTATTGAAGAGCAGAAATGTTATCTCTGTC	1434	2455	ATGCCACACAGTGGCCATTGGCCAAAAATTTACCTGAGAAATGAAAAATTAATGAACTGGCTTA	2514
QY	1276	ATTACTGAAACTCTCTAGAACTTTTACCTGAGTACTTGGACAGGAAACAATAAATTCAC	1335	2356	GAGAAATGCTATAAACAAGTGGCCACATTTTAAGAAACCAAGTGTATCAGGCCATGAGTT	2415
Db	1435	ATTACTGAAACCTCTAGAACTTTTACCTGAGTACTTGGACAGGAAACAATAAATTCAC	1494	2515	GAGAAATGCTATAAACAAGTGGCCACATTTTAAGAAACCAAGTGTATCAGGCCATGAGTT	2574
QY	1336	TTCCAGGTTATAGCCAGACAAATTTGGGAAGAGATATATGCAATATATGTCAGCTTAAG	1395	2416	TATGAATTAAGAGTGAATCCTGAAAGACCTTCAATATATGTTTATCATTATCTTCTAAA	2475
Db	1495	TTCCAGGTTATAGCCAGACAAATTTGGGAAGAGTATATGCAATATATGTCAGCTTAAG	1554	2575	TATGAATTAAGAGTGAATCCTGAAAGACCTTCAATATATGTTTATCATTATCTTCTAAA	2634
QY	1396	TATATCTGATCAGCAACCCACAAATATATGCAAGAAAGTAAAGATGCACTTCTTGAA	1455	2476	ATCCAGCATAGCAGGCTGAACATATGCAAGAAAGAGGAAACCAAGAAATTAAGAT	2535
Db	1555	TATATCTGATCAGCAACCCACAAATATATGCAAGAAAGTAAAGATGCACTTCTTGAA	1614	2635	ATCCAGCATAGCAGGCTGAACATATGCAAGAAAGAGGAAACCAAGAAATTAAGAT	2694
QY	1456	GGTTTTCGATCTTTTGAAGATTTCTTACCTGATGCAAGGGAATGGAAGAAATCCGAAGA	1515	2536	GAAGCATTTGCCGCCACCACCTCCCTGAAATTTCTGCCCTTTTTCAGCAAAAGTGAATAC	2595
Db	1615	GGTTTTCGATCTTTTGAAGATTTCTTACCTGATGCAAGGGAATGGAAGAAATCCGAAGA	1674	2695	GAAGCATTTGCCGCCACCACCTCCCTGAAATTTCTGCCCTTTTTCAGCAAAAGTGAATAC	2754
QY	1516	CAGTTTGGCAACACATTTGAAGTGGATCTGATTTGGGAGGCTGCCATTTCTATACAGATG	1575	2596	CTTCTCAACTGTGATATCATGATGATGATGATGATGATGATGATGATGATGATGATGATG	2655
Db	1675	CAAGTTGGACACACATTTGAAGTGGATCTGATTTGGGAGGCTGCCATTTCTATACAGATG	1734	2755	CTGCTCAGCTGTGATGTTATGATATATATATATATATATATATATATATATATATATAT	2814
QY	1576	CAATTTGAAGATATTTTACTCATGTTTCCAGAGTGGTGTGCTGTGATGAAAGACTCTTA	1635	2656	ACAGATTTCACTTGTGAGGAGGATGCTCCAAATGGCTTTTTCATATATTTTGGCATTTG	2715
Db	1735	CAACTTAAAGATATTTTGTCTCATGTTTCCAGAGTGGTGTGCTGTGATGAAAGACTCTTA	1794	2815	ACGGAGTCTAATCTGTGGACAGAGGATGCTGCAGATGGCTTTCCATATATTTGGCAGTG	2874
QY	1636	CTTGTGGCTTATAAGAAATGTCACAAAGCTGTGATGAGTGGATGAGTGGATGAGTGGATG	1695	2716	GGTTTACTAGAAGAGAGCAACAGCTTCAAAAAGCTCTGAAAGAGAGTGAATTTGAC	2775
Db	1795	CTTGTGGCTTATAAGAAATGTCACAAAGCTGTGATGAGTGGATGAGTGGATGAGTGGATG	1854	2875	GGCTTGTGGAAGAGAGCAGCAGCTTCAAGAAAGCTCTGAAAGAGAGTGGCTTTTTCAC	2934
QY	1696	AGTAGCAAGACAGTAGTACAAATCGTGTGACATAGTTTGAAGAAAGTCTTACAGAGTA	1755	2776	TTTTATATCAAGCTTCAAGATTTGGAAGTTTCAGCCAT - - - - - GAATATACAAATG	2826
Db	1855	AGTACCAAGACAGTAGTACAAATCGTGTGACATAGTTTGAAGAAAGTCTTACAGAGTA	1914	2935	TTTTATCAATCAAGCTTCAAGATTTGGAAGTTTCAGCCATGATGCTCAAGATATACAAATG	2994
QY	1756	TCTGAGGATCTTGAAGATATCATCTGCACTCTCTAGAGCCCTGCTGGTCTTCAATGTA	1815	2827	CTTTTGGAAAACATCAAGGAATTTCCCAAGTTTGAAGGCCAGAGGACATGATTAACGTGG	2886
Db	1915	TCTGAGGACCTTGTGAAGATATACCTGCACTCTCTAGAGCCCTGCTGGTCTTCAATGTA	1974	2995	CTCTTGGAAAGACTCAAGGAATCCCAATTTAGAAGGCCAGAGGACATGATTAACATGG	3054
QY	1816	CGTTTAAAGCAGCTGGTGTCTTTCAAGACTGCATGATTTGCTTTTTCAGGACTTT	1875	2887	ATACTTCAGATGTTTACACAGTGAAGCTTAAAGAGAAAAATCTGTTTATTTCTAGCA	2946
Db	1975	CGTTTAAAGCAGCTGGTGTCTTTCAAGACTGCATGATTTGCTTTTTCAGGACTTT	2034	3055	ATACTTCAGATGTTTACACAGTGAAGCTTAAAGAGAAAAATCTGTTTATTTAGTGGA	3114
QY	1876	CAAGTAGAGGTACTAGTGAATATCTTTACCTGTGCTGGTGTGCTGGTGTGCTGGTGTG	1935	2947	ACCACATCAGGATCGGAATCTTATTAAGAAATGATGAGATTTACTCATGATTAAGAAAAAGCA	3006
Db	2035	CAAGTAGAGGTACTAGTGAATATCTTTACCTGTGCTGGTGTGCTGGTGTGCTGGTGTG	2094	3115	ACCACATCAGGATCGGAATCTTATTAAGAAATGATGAGATTTACTCATGATTAAGAAAAAGCA	3174
QY	1936	GCTGAGATGTGGCGAAGAAATGGACTGTCTCTTATTAGCCAGGTGTTTATTACCAAGAT	1995	3007	GAACGAAAAAGAAAGCTGAAGCTGTAGGCTACATCGCCAGAGATCATGGCTCAGATG	3066
Db	2095	GCTGAGATGTGGCGAAGAAATGGACTGTCTCTTATTAGCCAGGTGTTTATTACCAAGAT	2154	3175	GAACGAAAAAGAAAGCTGAAGCTGTAGGCTTTCATCGCCAGAGATCATGGCCAGATG	3234
QY	1996	GTTAAGTCCAGAGAAAGATGATGATTAAGATATCATGCTTCAGATGTTGGTGATCT	2055	3067	TCCTGCCTTACAGAAAACTTCAATGAACTCAATACTCATGATGATGATGATGATGATGATG	3126
Db	2155	GTTAAGTCCAGAGAAAGATGATGATTAAGATATCATGCTTCAGATGTTGGTGATCT	2214	3235	TCCTGCCTTACAGAAAACTTCAATGAACTCAATACTCATGATGATGATGATGATGATGATG	3294
QY	2056	TTAATGGATCCCAATAGTCTTGTACTGTGTTTACCTGATGATGATGATGATGATGATG	2115	3127	ATGCTTGGGAAAGAAAGTTCATTTATGAGGAAAGAGCAGCCAGTCACTGATGATGATG	3186
Db	2215	TTAATGGATCCCAATAGTCTTGTACTGTGTTTACCTGATGATGATGATGATGATGATG	2274	3295	GTAACAGGAAAGAAAGTTCATTTATGAGGAAAGAGCAGCCAGTCACTGATGATGATG	3354
QY	2116	TTTAAAGACACATATCTACAAAGACAGGATTTGATTAACAAATATATATACATTAATA	2175	3187	TCTAGAAATTTGGGTCTTAAACAGGAGTCCATCTGTTTACTGAAAAAGAGGTGCTGACG	3246
Db	2275	TTTAAAGACACATATCTACAAAGACAGGATTTGATTAACAAATATATATACATTAATA	2334	3355	TCTAGAAATTTGGGTCTTAAACAGGAGTCCATCTGTTTACTGAAAAAGAGGTGCTGACG	3414
QY	2176	GAAGAAATGCTTCAGGCTCTCATCTATATTTGGGTGAGCGTTATGATACCTGGAGTGGGA	2235	3247	TGATCTCTTTGGCCAGAGAAACAGGAGTGAATAATAGAAATATATGATGATGATGATG	3306
Db	2335	GAAGAAATGCTTCAGGCTCTCATCTATATTTGGGTGAGCGTTATGATACCTGGAGTGGGA	2394	3415	TGATCTCTCTGCAAGAAAGAAACAGAGGTAAACTAGAAAAATATATGATGATGATG	3474
QY	2236	AATGTGACCAAGAGAGGTTCATATGATGAGAGAAATCATTCATGCTTGTGATTTGACCC	2295	3307	GCCTGTGCTCCAGAAATCTACTGCTTAAACAGCAGCAGGAGGAAAAACCCATGAACTCTCA	3366
Db	2395	AATGTGACCAAGAGAGGTTCATATGATGAGAGAAATCATTCATGCTTGTGATTTGACCC	2454	3475	GCATTTGTGAGAAATCCACCGCTTAAACCCAGCAGAGGAGGAGCCCTGTGGACCACTTA	3534
QY				3367	GGAGAGCCCTAGACCCACTTTTTCATGGATCCAGACTTTGGCATATGGAACCTTATACAGGA	3426



TITLE	Direct Submission	Qy	121	CATCATTTGGCACAATTTGGTCCAGAAATTTACTTGTCTGAATGAGACCCAGACTTGGAA	180
JOURNAL	Submitted (24-APR-1998) Division of Biology, 147-75, California	Db	235		
	Institute of Technology, 1200 E. California Bl., Pasadena, CA				
	91125, USA				
FEATURES	Location/Qualifiers	Qy	181	AAGCAGGAGAAAGTGTAACAATGTAATTTACTCCACTCGAATGGTACTATTGGGA	240
source	1. .6395	Db	295		
	/organism="Mus musculus"				
	/mol_type="mRNA"				
	/db_xref="taxon:10090"	Qy	241	GAAGATCCAGATATTTGCTTAGAGAAATTTGAACACAGTGGAGCAATTCAGCTTTGGGG	300
	/chromosome="2"	Db	355		
gene	/cell_line="MEL-C19"				
	1. .6395	Qy	301	AGGGTTTCAAAAGTGGAGACACACCTATTCTTTCAGGAGTATGCAATTCATCAACA	360
5'UTR	/gene="Ubr1"	Db	415		
	1. .114				
CDS	/gene="Ubr1"	Qy	361	TGCTACTCTCTATGGACTGCTTCCAGGACAGTGTTCATAAAATCATCTGTTACAAGATG	420
	115. .5388	Db	475		
	/note="UBR1p; recognition component of the N-end rule pathway"				
	/codon_start=1	Qy	421	CATACTTCTACTGGAGGAGGTTCTGTGACTGTGGAGACACAGAGGCATGGAAACTGGC	480
	/product="ubiquitin-protein ligase E3-alpha"	Db	535		
	/protein_id="AAC40165.1"				
	/db_xref="GI:3170887"	Qy	541	CCGTTGAATGAAGAGGTAATTTGTCAAGCCAGGAAATATTTCCCTCAGTGATAAATAT	600
	/translation="MADEMDGAERMDVPEPLAPORPASWMDQVDVYATFLHLHA	Db	655		
	QVPEYFAEMPDLEKQESVQMSILTPLEWLFGEEDPDICLEKLEKHSAPQLCKGV	Qy	601	GTCTAGAAATGACTATATGGGAGAGGAAAGAACTGCCCTCCCTGAATCCAGATAAGG	660
	FSGETTSCRCADIDPTCVLCMDFQSSVHKRHYKMTSTGGGDCDEKDEAWTKQ	Db	715	ATTCTAGAAATGACTATATGGGAGAGGAAAGAACTGCCCTCCCTGAATCCAGATAAGG	774
	PCVDHEPGRAITKESLCPNLEEVIAQARRIPPSVIKIVEMTWECEKELPPELQ	Qy	766	KNR-----	YVCVNDHSHYDH
	IREKNRYCVLFNDSVHSHVYLSIOLALDCELAELHTTAIDKGRRAVKAGVY	Db			
	ATCOAEKDIKSHENVSQHPHVEVLSVMAHAKFALRLGSMWMMKIMSYSDFRQI	Qy	775	GAGAAATGAACGATACATATTGTCTCTTTTCAACAGATGAGACCAATTCGTATGATCAT	834
	FCOACLVPEPGSNPCLLISRLMLWDAKLYKGAPKILHELIFSSFFEMEMKFLFAMEF	Db			
	VXYQLOKEYISDHERSISITALSVMQVTPVTLARHLIEQNVLSVITETLLEVL	Qy	835	GTGATCTACATCTGCAGAGAGCTCTAGATTGCGAGCTTGCAGAGGCAGCTGCACAG	894
	EYLDNRNKFQSGYSDKLRGVAVICDLKYILISKPVITWTERLRAQFLEGRSFLKI	Db			
	LTCMGMERIRFQSGHIEVDFPDAAIAIOMQLKNILMFQWACDEDLVVAYKE	Qy	736	ACTGCCATTTGCAAGAGGGTCTGGGCTGTGTAAGGGGGGAGCTTATCTGCTGCCAG	795
	CHKAVMCTSNFQSTTQVQCGHSLKTSKVSDELVSILPLSLTLAGLHVRLSR	Db	895	ACTGCCATTCGACAAAGAGGGTCTGGGCTGTCAAAGCAGGTGTGTATGCCACTTGCAG	954
	LGAIRLFEVDFQSVLEVYPLRCLVLAOVVAEMWRRNGLSLISOVFYYQDVK	Qy	796	GAACCAAGGAAGATATAAGAGCTATTCAAGAAATGTCTCTCAACATCCATTCATGTA	855
	CREMYDKDIIMLOIGASTMDPNKFLVLLVRYELVDANKTISTQDQLIKOYNLI	Db	955	GAACCAAGGAGGATATAAGAGCTACTCAGAGAAACGCTCTCTCAGCACCCCTCCATGTG	1014
	EMLOVLYIGERYVPGVNTREYVIRETHILICIEPMPHSAIARNLPENNET	Qy	856	GAGCTATTACACTCAGAGATTATGCTCATCAGAAATTTGCTTCCGCTCTGCTGCTCGG	915
	GLENVINKATKPCGVSGVVELKDSKDFNYFIYHSKTOHKAEMQKRRKQ	Db	1015		
	ENKDALPPPPPEPCAFSVKVNLSQDMYILTFIFRAYDVTESNLWTEGLQMA	Qy	916	ATGAACAAATTTATGAGCTATTCAAGTGACATTTAGGCAGATCTTTTCCCAAGCATGCCCT	975
	PHILALLLEQQQLQKAPKEEVAFDFYHKASGSSAMNAQNIOMLLERLKGIPOLE	Db	1075	ATGAACAAATTTATGAGCTATTCAAGTGACATTTAGGCAGATATTTCAGCAGGCTGCCTC	1134
	GKDMITWLOMFTDKRLREKSLVATTSGLECIKSEIITHDKBAERKRAEAR	Qy	976	AGAGAAGACCTGACTCGGAGAAATCCCTGTCTCATAGCAGGTTTAATGCTTTGGGATGCA	1035
	LHROKIMAOASLQKNFIETHLMYDNTSEVTKEDSIMEESTSAVSEARLQPK	Db	1135	GTAGAAGAACCTGGCTCTGAAATCCCTGTCTTATAAGCAGACTAATGCTTTGGGATGCA	1194
	RGPATKEVLTCLICQREQVYKLENNAMVLSNCVQKSTALTOHGRKGVDDLGETDP	Qy	1036	AAGCTTTTATAAGGTGCCCGTAAAGTCCCTCATGAAATTTGTAAGTTCAGCAGATTTTTTATG	1095
	LFMDPLDGTGTCGGVHVAWQYFEAVOLSSQRIHVLDLSEGEYLCLPCK	Db	1195	AACTTTTATAAGGTGCCCGTAAAGTCCCTCATGAAATTTGTAAGTTCAGTATTTTATG	1254
	SLCNTVPIPILOPKINSENAEALQTLTARWQTLVRLATSGYNIHAKGAPAPV	Qy	1096	GAGATGAATACAAAAAATCTTTTGTCTATGAAATTTGTGAAGTATTATAAACAACTGCAG	1155
	VLFNMGDSTFEHLSIFGVQSSVYKSIKEMVILFATTIYRIGLKVPPDELDPR	Db	1255	GAGATGAATACAAAAAATCTTTTGTCTATGAAATTTGTGAAGTATTATAAACAACTGCAG	1314
	VPMWTSTCAFTIQAENLLGDGKPLFGALONROHSLKALMOFAVAORATCPVLI	Qy			
	HKHLARLLSLPNLQSENTPGLLSVDLFHLVGLAVLAPPSLYWDDTVLOPSPSS	Db			
	YNHLYFLHITMAHMLQILLTDTDLSPGPPARGEDSEERACSAFFVEVSQHTDG	Qy			
	LFCCGAPGWLMSLRNGITPYLRCAALLFHYLLGVAPPEELFANSAEGEFSALCSTYL	Db			
	SLPTNLLFLFYWDPTIRLLQRCWGDPAALLSLKOKSAVVYPRKNSLIPELDEYS	Qy			
	CLLNQASHPCRSADDERKHPVLCFLCGAILCSQNICQEIYVNGEVEGACVPHLHC	Db			
	GAGVCIFLKRRCRVVLRGKARGCAYAPYLDEYGETDPLGRGNPLHLSRERYKL	Qy			
	HLVWQHQHIIIEIARSQETNQMLFGFNWOLL"	Db			
	5389. .6395				
3'UTR	/gene="Ubr1"	Qy			
	5429. .5434	Db			
polya_site	/gene="Ubr1"				
	1802 a 1519 g 1698 t				
BASE COUNT					
ORIGIN					
	Query Match 78.5%; Score 4085.6; DB 10; Length 6395;				
	Best Local Similarity 86.6%; Pred. No. 0;				
	Matches 4567; Conservative 9; Mismatches 629; Indels 69; Gaps 4;				
Qy	1 ATGGCGGACGAGAGGCTGGAGTACTGAGAGGATGGAATACAGCGCGGAGTACCCCGAG				
Db	115 ATGGCGGACGAGAGATGACGCGCGGAGAGGATGAGCTAGCGCGGAGCGCTCCCTG				
Qy	61 ACCCCTCAGCGTGTGGTGGGATCAGCAAGTTGATTTTATCTACTGCTTCTTG				
Db	175 GCCCGCAGCGCGGCGATCGTGGTGGGATCAGCAAGTTGATTTTATCTACTGCTTCTTA				

[illegible]



QY	3367	GGAGAA	CCCTAGAC	CCCACTTTTCATGGATCCAGACTTGGCATATGGAACTTTATACAGGA	3426
Db					
QY	3355	GGGAAA	ACATGGAC	CCCTCTTTTCATGGATCCAGACTTGGCATATGGAACTTTATACAGGA	3594
Db					
QY	3427	AGCTGTGGT	CATGTAATGCAGCAGCTGTGCTGGCAGAAAGTATTTTGAAGCTGTACAGCTG	3486	
Db					
QY	3395	AGCTGTGGT	CATGTAATGCATGCAGTGTGCTGGCAGAAAGTATTTTGAAGCTGTGCAGCTG	3654	
Db					
QY	3487	AGCTCTCAG	AGCGCATTCATGTGTGACCTTTTGTGACTTGGAAAGTGAGAAATATCTTTTGC	3546	
Db					
QY	3655	AGCTCGCAG	CGCATTCACGTAGACCTGTTTGACCTGGAGAGCGCGAGTATACG	3714	
Db					
QY	3547	CCCTCTTGC	AAAATCTCTGTGCAATCTGATCCCAATTTATTTCCCTTGGCAACTCAAAG	3606	
Db					
QY	3715	CCGCTCTGC	AAAGTCTCTCTCAACACTGTATCCCATCATCCCTTTTGCAGCCGCAAG	3774	
Db					
QY	3607	ATAAACAGT	CAGAAATGCAGATGCTCTGTGCTCAACTTTTGACCTTGGCCAGCTGGATACAG	3666	
Db					
QY	3775	ATCAACAGT	CAGAAATGCGGAGGCTCTGTGCTCAACTTTTGACCTTGGCCCGGTGGATACAG	3834	
Db					
QY	3667	ACTGTCTGC	CCAGAAATATCAGGTTATAATATAAGCATGCTAAAGAGAGAAAACC---	3723	
Db					
QY	3835	ACTGTCTCT	GCAGAAATATCGGTTATAATATAAAGCATGCTAAAGAGAGAAAGCCAGCA	3894	
Db					
QY	3724	ATTTCCAT	TTTCTTTAATCAAGGAATGGGAGATTTCTACTTTGGAGTTCCATTTCCATCCTG	3783	
Db					
QY	3895	GTTCCCTGT	CTGTGTTAATCAAGGAATGGGAGATTTCAACTTTTGAGTTTCAATTTCCATCCTG	3954	
Db					
QY	3784	AGTTTTGGC	GTCTGAGTCTCGAATTAATATTCAAATAGCATCAAGGAATGGTTATTTCTC	3843	
Db					
QY	3955	AGTTTTGGG	AGTTCAGTCTCGGTGAATATTCAAATAGTATCAAGGAATGGTTATTTCTC	4014	
Db					
QY	3844	TTTTGCCA	CAACAATTTATAGAATTTGATTTGAAAGTGCCACTGATGAAAGGGATCTCTCGA	3903	
Db					
QY	4015	TTGCCACA	CAACAATTTACGAATTTGGCTGAAAGTGCTCTCTGATGAATAGACCCACGA	4074	
Db					
QY	3904	GTCCCATG	CTGACCTGGACACCTGGCTTTTCACTATCCAGGCAATTTGAAATCTATTTG	3963	
Db					
QY	4075	GTGCCATG	ATGACCTGGACACCTGGCTTTTCACTATCCACCTGCAAGCAATTCGAAACCTGTTG	4134	
Db					
QY	3964	GGAGATGA	AGGAAAACTCTGTTTGGAGCACTTCAAAATAGGACGATATGCTCTCAAA	4023	
Db					
QY	4135	GGAGATGA	AGGAAAACTCTATTTGGAGCACTTCAAAATAGACGATACGCGTTCTGAAG	4194	
Db					
QY	4024	GCATTAAT	GCAGTTTGCAGTTGCACAGAGATTAACCTGTCTCCTCAGGTCCTGATACAGAAA	4083	
Db					
QY	4195	GCGCTAAT	GCAGTTTGCAGTTGCACAGAGGGCTACCTGCGCTCAGGTCCTGATACACAAA	4254	
Db					
QY	4084	CATCTGGT	TCGCTCTTCATCAGTTGTTCCTTAAACATAAAATCAGAAAGTACACCATGC	4143	
Db					
QY	4255	CATCTGGC	TCGCTCTTCAGTTATTTCTTCTTAACTGCAATCAGAAAAATACACAGGC	4314	
Db					
QY	4144	CTTCTGTC	TATAGATCTGTTTCAATGTTTGGTGGGTGCTGTGTAGCATTTCCATCCTGTTG	4203	
Db					
QY	4315	CTTCTGTC	TCTGATCTCTTCCATGTTTCTGGTGGGCGAGTCTTAGCGTTTCCCATCCTGTTG	4374	
Db					
QY	4204	TATTTGGAT	CACCCCTGTGATCTGCAGCTTCTTCAGTTAGTTTCTTCTCTATACCACTTT	4263	
Db					
QY	4375	TATTTGGAT	GACACCGTGGATCTGCAGCGCTGCGCACTTAGTTTCTTCATATACCACTTC	4434	
Db					
QY	4264	TATCTCTTC	CATTTGATCAGCATGGCACATGCTTCAGATACTACTTACATACAGTAGACA--	4321	
Db					
QY	4435	TATCTCTTC	ATCTGATCACCATGGCGCATGCTTCAGATCTCTCTTCAACAGATACA	4494	
Db					
QY	4322	-----	-----CAGCCCTACCCCTGTGCTCAGGTTTCAAGAGACAGTGAAGAGGCTCATTC	4371	
Db					
QY	4495	GATCTGTCT	CCAGGGCGCGCTTGCTGAGGGTGAAGAGGATAGTGAGGAGGCTCGCTGT	4554	
Db					
QY	4372	GCATCTCT	TTCTTTTGACAGAAATTTCTCAATATACAGTGGCTCCATTTGGGTGTGATATT	4431	
Db					
QY	4555	GCATCTCT	TTCTTTTGTGGAAGTGTGCGACACAGACGCGCCTCACTGCGGTGGGTGCT	4614	
Db					

QY	4432	CTTGGCTGGTATT	TTGTGGGCTCTC	ACTGAAGAATGG	CATCACCCCTTAT	CTTCTGGCTGTGCT	4491
Db	4615	CCCGGGTGGTAC	TGTGGCTCTCC	CTGAGAAACGG	CATCACCCCTTAC	CTTACCTCCGCTGTGCT	4674
QY	4492	GCATTGTTTTT	CCACTATTAT	TACTTGGGGTAA	CTCCGCTTGA	GGAACATGCATACCAATTC	4551
Db	4675	GCATGCTTTT	CCACTATTAT	TACTTGGAGTAG	CTCCGCTTGA	GGAACATGTTTGCCATTC	4734
QY	4552	GCAGAGGAGAG	TACAGTGC	ACTCTCTAG	ACTATCTTTT	AGCTACAAATTTGTTCC	4611
Db	4735	GCTGAAGAGAA	TTTCAGTGC	ACTCTCTAG	CTATCTTTT	TACCCACAAATTTGTTCC	4794
QY	4612	CTCTTCCAGAA	TATTTGGGAT	CTGTAAAGCCCTT	CTCTCCAGAGCG	GTGTCAGATCCT	4671
Db	4795	CTTTTCCAGAA	TATTTGGGAT	CTGTAAAGCCCTT	CTCTACAGAGG	TGGTGTGGAGATCCT	4854
QY	4672	GCCTTACTAA	CTGTTTGA	AGCAAAAAA	CACCGTGTG	CAGTACCCCTAGAAAAGAAAT	4731
Db	4855	GCCTTACTCA	GTCTTTGA	AGCAGAAAG	TGCTGTGCT	CAGTACCCCTAGAAAAGAAAT	4914
QY	4732	AGTTTGATAG	AGCTTCTCT	GATGACTAT	AGTGCCTCCT	GAATCAAGCTTCTCATTTTCAGG	4791
Db	4915	AGTTTGATAG	AGCTTCTCT	GAGGACTAC	AGTGTCTTAA	ATCATCAGGCTTCTCACTTTAGG	4974
QY	4792	TGCCACGGT	CTCCAGAT	GATGAGCA	AAACATCCTCT	GTGCGCTTTTCTGTGGGGCT	4851
Db	4975	TGTCACGGT	CTCCAGAT	GATGAGCA	AAACATCCTCT	GTGCGCTTTTCTGTGGGGCT	5034
QY	4852	ATACTATGTT	CTCAGAAC	ATTTCCTG	CCAGGAAAT	TGTGAACGGGAAGAGGTTGAGCT	4911
Db	5035	ATCCTGTGTT	CTCAGAAC	ATCTGTT	CCAGAAAT	AGTGAATGGGAAGAGGTTGAGCG	5094
QY	4912	TGCATTTTTC	CACGCACTT	CACCTGTG	GAGCGGAGT	CTGCATTTTCTTAAAAATCAGAGAA	4971
Db	5095	TGCGTTTTC	CATCGCTT	CAATTTG	TGCTGGAGT	CTGCATTTTCTTAAAAATCAGAGAA	5154
QY	4972	TGCCGAGT	GTGCTGTT	GAAGTAA	AGCCAGAGG	CTGTGCGCTATCCAGCTCTTACTTTG	5031
Db	5155	TGCAGGGT	GTGCTGTT	GAAGAAAG	CCAGAGGCTGT	GCTTACCCAGCCCTTACTTTG	5214
QY	5032	GATGAATAT	GGAGAAAC	AGACCTTG	CGCTTGA	AGAGGGCAACCCCTTCATTTATCTCGT	5091
Db	5215	GATGAATAT	GGAGAAAC	AGACCCAGG	CTTAA	AGAGAGGAAACCCCTTCATTTATCTCGG	5274
QY	5092	GAGCGGTAT	CGGAAGCT	TCCATTTT	GGTCTG	GCAACACACTGCATTTATAGAAGAGATTGCT	5151
Db	5275	GAGCGGTAT	CGGAAGCT	GCATTTT	GGTCTG	GCAACACACTGCATTTATAGAAGAGATTGCT	5334
QY	5152	AGGAGCCAG	AGAGACTA	ATCAGATG	TATTTTGG	ATTCACTGGCAGTTACTGTGA	5205
Db	5335	CGGAGCCAG	GAGAGACTA	ATCAGATG	CTATT	TGGATTTAACTGGCAGTTACTGTGA	5388
RESULT 6							
AX714232				3059 bp	DNA	linear	PAT 15-APR-2003
LOCUS				Sequence 916 from Patent EP1293569.			
DEFINITION							
ACCESSION				AX714232			
VERSION				AX714232.1		GI-29889184	
KEYWORDS							
SOURCE				Homo sapiens (human)			
ORGANISM				Homo sapiens			
				Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
				Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
REFERENCE				1			
AUTHORS				Isogai, T., Sugiyama, T., Otsuki, T., Wakamatsu, A., Sato, H., Ishii, S., Yamamoto, J. I., Isono, Y., Hio, Y., Otsuka, K., Nagai, K., Irie, R., Tamechika, I., Seki, N., Yoshikawa, T., Otsuka, M., Nagahari, K. and Masubo, Y.			
TITLE				Full-length cDNAs			
JOURNAL				Patent: EP 1293569-A 916 19-MAR-2003;			
				Helix Research Institute (JP) ; Research Association for Biotechnology (JP)			



FEATURES		Location/Qualifiers	
source		1..3059	
		/organism="Homo sapiens"	
		/mol_type="genomic DNA"	
		/db_xref="taxon:9606"	
BASE COUNT	969 a	546 c	694 g 850 t
ORIGIN			
Query Match 56.2%; Score 2927.4; DB 6; Length 3059;			
Best Local Similarity 98.0%; Pred. No. 0;			
Matches 2981; Conservative 10; Mismatches 7; Indels 45; Gaps 1;			
QY	1	ATGCGGACAGAGAGGCTGGAGTACTGAGAGATGAAATCAGCGGGAGTTACCCAG	60
Db	17	ATGCGGACAGAGAGGCTGGAGTACTGAGGGATGAAATCAGCGGGAGTTACCCAG	76
QY	61	ACCCCTCAGCGTCTGCATCTTGGTGGGATCACAAGTTCATTTTATACGTCTTCTTG	120
Db	77	ACCCCTCAGCGTCTGCATCTTGGTGGGATCACAAGTTCATTTTATACGTCTTCTTG	136
QY	121	CATCATTTGGCACAATTTGGTGGCAGAAATTTACTTTGCTGAAATGGACCCAGACTTGGAA	180
Db	137	CATCATTTGGCACAATTTGGTGGCAGAAATTTACTTTGCTGAAATGGACCCAGACTTGGAA	196
QY	181	AAGCAGAGAAAGTGTACAAATGTCAAAATGTCAAATTCACCTCCACTGGATGGTACTTATTTGGA	240
Db	197	AAGCAGAGAAAGTGTACAAATGTCAAAATGTCAAATTCACCTCCACTGGATGGTACTTATTTGGA	256
QY	241	GAAGTCCAGATATTTGCTTAGAGAAATGAAGCACAGTGGAGCATTTTTCAGCTTTGGGG	300
Db	257	GAAGTCCAGATATTTGCTTAGAGAAATGAAGCACAGTGGAGCATTTTTCAGCTTTGGGG	316
QY	301	AGGGTTTTCAAAGTGGAGAGACAACTATTCTTGCAAGGATTTGTCATTTGATTCACAACA	360
Db	317	AGGGTTTTCAAAGTGGAGAGACAACTATTCTTGCAAGGATTTGTCATTTGATTCACAACA	376
QY	361	TGTGTACTGTATGAGTCTGCTCCAGGACAGTGTTCATAAAATCATCGTTACAAGATG	420
Db	377	TGTGTACTGTATGAGTCTGCTCCAGGACAGTGTTCATAAAATCATCGTTACAAGATG	436
QY	421	CATACCTTCTACTGGAGAGGGTTCTGTGACTGTGGAGACACAGAGCATCGGAAACTGGC	480
Db	437	CATACCTTCTACTGGAGAGGGTTCTGTGACTGTGGAGACACAGAGCATCGGAAACTGGC	496
QY	481	CGTTTTTGTGTAATCATGAACCTGGAAGAGAGTACTATATAAGAGAAATTCACGCTGT	540
Db	497	CGTTTTTGTGTAATCATGAACCTGGAAGAGAGTACTATATAAGAGAAATTCACGCTGT	556
QY	541	CGTTTGAATGAAGAGTAAATTTGCCAAGCCAGGAAATATTTTCTTCACTGATATAAATAT	600
Db	557	CGTTTGAATGAAGAGTAAATTTGCCAAGCCAGGAAATATTTTCTTCACTGATATAAATAT	616
QY	601	GTCGTAGAAATGACATATATGGGAAGAGGAAAGAACTGCCCTCGAACTCCAGATAAGG	660
Db	617	GTCGTAGAAATGACATATATGGGAAGAGGAAAGAACTGCCCTCGAACTCCAGATAAGG	676
QY	661	KNR-----YVCVNDHHSYDH 675	
Db	677	GAGAAAATGAAGATACTATTTGTCTCTTTTCAATGATGAACACCATTTTCATATGACCAC	
QY	676	GTCAATATACAGCCTACAAAGAGCTCTTGACTGTGAGCTGCGAGAGCCGAGTTGCATACC	735
Db	737	GTCAATATACAGCCTACAAAGAGCTCTTGACTGTGAGCTGCGAGAGCCGAGTTGCATACC	796
QY	736	ACTGCCATTTGACAAAGAGGGTCTGCGGCTGTGTTAAAGGGGAGCTTATGCTCTTGGCCAG	795
Db	797	ACTGCCATTTGACAAAGAGGGTCTGCGGCTGTGTTAAAGGGGAGCTTATGCTCTTGGCCAG	856
QY	796	GAAGCAAGGAAGATATAAAGAGTCATTCAGAAATGTCTCTCAACATCCACTTCATGTA	855
Db	857	GAAGCAAGGAAGATATAAAGAGTCATTCAGAAATGTCTCTCAACATCCACTTCATGTA	916
QY	856	GAAGTATTACACTCAGAGATTATGGCTCATCAGAAATTTGCTTTGCGCTTGGTTCCTG	915
Db	917	GAAGTATTACACTCAGAGATTATGGCTCATCAGAAATTTGCTTTGCGCTTGGTTCCTG	976
QY	916	ATGAACAAAATATAGAGCTATTTCAAGTACTTTAGGCAGATCTTTTTCGAAGCATGCCCT	975
Db	977	ATGAACAAAATATAGAGCTATTTCAAGTACTTTAGGCAGATCTTTTTCGAAGCATGCCCT	1036
QY	976	AGAGAAGAACCTGACTCGGAGAAATCCCTGTCTCATAGCAGGTTAATGCTTTGGGATGCA	1035
Db	1037	AGAGAAGAACCTGACTCGGAGAAATCCCTGTCTCATAGCAGGTTAATGCTTTGGGATGCA	1096
QY	1036	AAGCTTTATAAAGGTGCCCGTAAGATCTTCATGAATTCATCTTCAGCAGTTTTCATG	1095
Db	1097	AAGCTTTATAAAGGTGCCCGTAAGATCTTCATGAATTCATCTTCAGCAGTTTTCATG	1156
QY	1096	GAGATGAATACAAAAAATCTTTTCTATGAAATTTGTGAAGTATTTATAACAACCTGCAG	1155
Db	1157	GAGATGAATACAAAAAATCTTTTCTATGAAATTTGTGAAGTATTTATAACAACCTGCAG	1216
QY	1156	AAAGAATATATCAGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	1215
Db	1217	AAAGAATATATCAGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	1276
QY	1216	ATGTTTACTGTTCTTACTCTGGCTCGACATCTTATGAAGACGAGATGTTATCTCTGTC	1275
Db	1277	ATGTTTACTGTTCTTACTCTGGCTCGACATCTTATGAAGACGAGATGTTATCTCTGTC	1336
QY	1276	ATTACTGAAATCTGCTAGAAATTTTACCTCAGTACTTTGGACAGGAAACAATAAATTCAC	1335
Db	1337	ATTACTGAAATCTGCTAGAAATTTTACCTCAGTACTTTGGACAGGAAACAATAAATTCAC	1396
QY	1336	TTCCAGGGTTATAGCCAGGACAAATGGGAAGAGTATATGCAAGTAAATGTGACCTAAAG	1395
Db	1397	TTCCAGGGTTATAGCCAGGACAAATGGGAAGAGTATATGCAAGTAAATGTGACCTAAAG	1456
QY	1396	TATATCTGTATCAGCAACCCACAATATGGACAAAGATTTAAGAAATGCACTTCTTGAA	1455
Db	1457	TATATCTGTATCAGCAACCCACAATATGGACAAAGATTTAAGAAATGCACTTCTTGAA	1516
QY	1456	GGTTTTGCACTTTTTTGAAGATTTTACCTGATGAGGGAATGGGAAGATTCGGAAGA	1515
Db	1517	GGTTTTGCACTTTTTTGAAGATTTTACCTGATGAGGGAATGGGAAGATTCGGAAGA	1576
QY	1516	CAGTTGGGCAACATATGAAATGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGA	1575
Db	1577	CAGTTGGGCAACATATGAAATGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGA	1636
QY	1576	CAATTGAAGAAATATTTTACTCATCTGTCAGAGTGGTGTGCTTGTGATGAAGAACTCTTA	1635
Db	1637	CAATTGAAGAAATATTTTACTCATCTGTCAGAGTGGTGTGCTTGTGATGAAGAACTCTTA	1696
QY	1636	CTTTGGCTTTATAAAGAAATGTCACAAAGCTGTGATGAGGTGCAGTACCAGTTTCATATCT	1695
Db	1697	CTTTGGCTTTATAAAGAAATGTCACAAAGCTGTGATGAGGTGCAGTACCAGTTTCATATCT	1756
QY	1696	AGTAGCAAGACAGTAGTACAAATCGTGTGGACATAGTTTGGAAACAAAGTCCCTACAGAGTA	1755
Db	1757	AGTAGCAAGACAGTAGTACAAATCGTGTGGACATAGTTTGGAAACAAAGTCCCTACAGAGTA	1816
QY	1756	TCGTGAGATCTTGTAGCATACATCTGCCACTCTCTAGGACCCCTTGTGCTTTCATATGTA	1815
Db	1817	TCGTGAGATCTTGTAGCATACATCTGCCACTCTCTAGGACCCCTTGTGCTTTCATATGTA	1876
QY	1816	CGTTTTAAGCAGGCTGGTGTCTTCAAGACTGCATCAATTTGCTCTTTTGAGGACTTTT	1875
Db	1877	CGTTTTAAGCAGGCTGGTGTCTTCAAGACTGCATCAATTTGCTCTTTTGAGGACTTTT	1936
QY	1876	CAAGTAGAGTACTAGTGGAAATCCCTTTACGTTGTCTGCTGTTGTTGCTGCTGCTGCTG	1935
Db	1937	CAAGTAGAGTACTAGTGGAAATCCCTTTACGTTGTCTGCTGTTGTTGCTGCTGCTGCTG	1996
QY	1936	GCTGAGATGTGGGGAAGAAATGGACGTCTCTTTATAGCCAGGTTTTCATACCAAGAT	1995

Db 1997 GCTGAGATGTGGCGAAGAAATGACTGCTCTCTTATTAGCCAGGTGTTTATTATACCAAGAT 2056  
QY 1996 GTTAAAGTGCAGAGAAGAAATGTATGATAAAGATATATCATCATGCTTCAGATTGGTGCATCT 2055  
Db 2057 GTTAAAGTGCAGAGAAGAAATGTATGATAAAGATATATCATCATGCTTCAGATTGGTGCATCT 2116  
QY 2056 TTAATGGATCCCAATAAGTCTTGTGTACTGTGTTACTTACAGAGGTATGAACCTTGCAGAGCT 2115  
Db 2117 TTAATGGATCCCAATAAGTCTTGTGTACTGTGTTACTTACAGAGGTATGAACCTTGCAGAGCT 2176  
QY 2116 TTTAACAAGACCATATCTCAAAAGACCAAGGATTTGATTAAACAATATATACACTAATA 2175  
Db 2177 TTTAACAAGACCATATCTCAAAAGACCAAGGATTTGATTAAACAATATATACACTAATA 2236  
QY 2176 GAAGAAATGCTTCAGTCTCATCTATATTTGGGTGAGCGTTATGTACCTGGAGTGGGA 2235  
Db 2237 GAAGAAATGCTTCAGTCTCATCTATATTTGGGTGAGCGTTATGTACCTGGAGTGGGA 2296  
QY 2236 AATGTGACCAAAAGAGAGGTCACAATGAGAGAAATCATTCACATTCGCTTTCATTGAACCC 2295  
Db 2297 AATGTGACCAAAAGAGAGGTCACAATGAGAGAAATCATTCACATTCGCTTTCATTGAACCC 2356  
QY 2296 ATGCCACACAGTGCCATTGCCAAAATTTACCTGAGAATGAAAATATGAACCTGGCTTA 2355  
Db 2357 ATGCCACACAGTGCCATTGCCAAAATTTACCTGAGAATGAAAATATGAACCTGGCTTA 2416  
QY 2356 GAGATGTGATTAACAAGTGGCCACATTTAAGAACCGAGTGTATCAGGCCATGGAGTT 2415  
Db 2417 GAGATGTGATTAACAAGTGGCCACATTTAAGAACCGAGTGTATCAGGCCATGGAGTT 2476  
QY 2416 TATGAACCTAAAGATGAATCACTGAAGACTTCAATATGTACTTTTATCATTTACTCCAAA 2475  
Db 2477 TATGAACCTAAAGATGAATCACTGAAGACTTCAATATGTACTTTTATCATTTACTCCAAA 2536  
QY 2476 ACCCAGCATGACCAAGGCTGAACATATGCAGAGAAAGGAGAGAAACAAAGAAACAAAGAT 2535  
Db 2537 ACCCAGCATGACCAAGGCTGAACATATGCAGAGAAAGGAGAGAAACAAAGAAACAAAGAT 2596  
QY 2536 GAAGCATTCGCCGCCACCACCTCCTGAATTTCTGCCTTTCAGCAAAAGTGAATAC 2595  
Db 2597 GAAGCATTCGCCGCCACCACCTCCTGAATTTCTGCCTTTCAGCAAAAGTGAATAC 2656  
QY 2596 CTTCTCAACTGTGATATCATGATGTACATTCAGGACCGTATTTAGCGGGCAATAGAC 2655  
Db 2657 CTTCTCAACTGTGATATCATGATGTACATTCAGGACCGTATTTAGCGGGCAATAGAC 2716  
QY 2656 ACAGATTCTTAACCTGTGGACCGAAGGATGCTCCAAATGCTTTTCATATTCGGCATTTG 2715  
Db 2717 ACAGATTCTTAACCTGTGGACCGAAGGATGCTCCAAATGCTTTTCATATTCGGCATTTG 2776  
QY 2716 GCTTTACTAGAGAGAAGCAACAGCTTCAAAAGCTCCTGAAGAAAGTAAATACATTTGAC 2775  
Db 2777 GCTTTACTAGAGAGAAGCAACAGCTTCAAAAGCTCCTGAAGAAAGTAAATACATTTGAC 2836  
QY 2776 TTTTATCATAGAGCTTCAAGATTTGGAAGTTCAGCCATGAATATACAAATGCTTTTGGAA 2835  
Db 2837 TTTTATCATAGAGCTTCAAGATTTGGAAGTTCAGCCATGAATATACAAATGCTTTTGGAA 2896  
QY 2836 AACTCAAGAGAAATTCCTCCAGTTAGAGCCAGAGGACATGATACGTGGATACCTCAG 2895  
Db 2897 AACTCAAGAGAAATTCCTCCAGTTAGAGCCAGAGGACATGATACGTGGATACCTCAG 2956  
QY 2896 ATGTTTGACACAGTGAAGGATTAAGAGAAAATCTTGTAAATCTAGCAACCAATCA 2955  
Db 2957 ATGTTTGACACAGTGAAGGATTAAGAGAAAATCTTGTAAATCTAGCAACCAATCA 3016  
QY 2956 GGATCGGAATCTATTAAAGATGATGAGATTACTCATGATAAAG 2998  
Db 3017 GGATCGGAATCTATTAAAGATGATGAGATTACTCATGATAAAG 3059

RESULT 7  
AK056441

LOCUS AK056441 3059 bp mRNA linear PRI 01-AUG-2002  
DEFINITION Homo sapiens CDNA FLJ31879 fis, clone NT2RP7002619, highly similar  
to Mus musculus ubiquitin-protein ligase E3-alpha (Ubr1) mRNA.  
ACCESSION AK056441  
VERSION AK056441.1 GI:16551844  
KEYWORDS oligo capping; fis (full insert sequence).  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE 1  
AUTHORS Tashiro,H., Yamazaki,M., Watanabe,K., Kumagai,A., Itakura,S.,  
Fukuzumi,Y., Fujimori,Y., Komiyama,M., Sugiyama,T., Irie,R.,  
Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y.,  
Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Yamashita,H.,  
Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K.,  
Wagatsuma,M., Murakawa,K., Kanehori,K., Takahashi-Fujii,A.,  
Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S.,  
Nagahara,K., Masuho,Y., Nagai,K. and Isogai,T.  
NEDO human cDNA sequencing project  
Unpublished  
2 (bases 1 to 3059)  
REFERENCE Isogai,T., Otsuki,T. and Sugiyama,T.  
Direct Submission  
Submitted (24-OCT-2001) Takao Isogai, Helix Research Institute,  
Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
(E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)  
NEDO human cDNA sequencing project supported by Ministry of  
Economy, Trade and Industry of Japan; cDNA full insert sequencing:  
Research Association for Biotechnology (RAB); cDNA library  
construction: Helix Research Institute (HRI) (supported by Japan  
Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,  
HRI, and Biotechnology Center, National Institute of Technology and  
Evaluation; clone selection for full insert sequencing: RAB and  
HRI.  
FEATURES  
source Location/Qualifiers  
1. .3059  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="NT2RP7002619"  
/cell\_line="NT2"  
/cell\_type="teratocarcinoma"  
/clone\_lib="NT2RP7"  
/note="cloning vector: pME18SFL3-mRNA from NT2 neuronal  
precursor cells after 5-weeks retinoic acid (RA)  
induction."  
BASE COUNT 969 a 546 c 694 g 850 t  
ORIGIN  
Query Match 56.2%; Score 2927.4; DB 9; Length 3059;  
Best Local Similarity 98.0%; Pred. No. 0;  
Matches 2981; Conservative 10; Mismatches 7; Indels 45; Gaps 1;  
QY 1 ATGCGGACGACGAGGCTGGAGGTACTGAGAGGATGGAAATCAGCCGGAGTTACCCAG 60  
Db 17 ATGCGGACGACGAGGCTGGAGGTACTGAGGAGTGGAAATCAGCCGGAGTTACCCAG 76  
QY 61 ACCCTCAGGCTCGGCATCTTGTGGATCAGCAAGTTGATTTTATCTACTGCTTTCTTG 120  
Db 77 ACCCTCAGGCTCTGGCATCTTGTGGATCAGCAAGTTGATTTTATCTACTGCTTTCTTG 136  
QY 121 CATCATTTGGCACAAATTTGGTCCCAAAATTTTCTGCTGAATGACCCAGACTTGGAA 180  
Db 137 CATCATTTGGCACAAATTTGGTCCCAAAATTTTCTGCTGAATGACCCAGACTTGGAA 196  
QY 181 AAGCAGGAGGAAAGTGTACAATGTCAATATTTCACTCCACTGGAATGGTACTTATTTGA 240  
Db 197 AAGCAGGAGGAAAGTGTACAATGTCAATATTTCACTCCACTGGAATGGTACTTATTTGA 256  
QY 241 GAAGATCCAGATATTTGCTTTAGAGAAATTTGAAGCACAGTGGAGCAATTCAGCTTTGTGG 300  
Db 257 GAAGATCCAGATATTTGCTTTAGAGAAATTTGAAGCACAGTGGAGCAATTCAGCTTTGTGG 316

QY 301 AGGTTTTCAAAAGTGGAGACACACCTATTCTTCCAGGGATTGTGCAATTCATCCAAACA 360  
DB 317 AGGTTTTCAAAAGTGGAGACACACCTATTCTTCCAGGGATTGTGCAATTCATCCAAACA 376  
QY 361 TGTACTCTGTATGGACTGCTCCAGACACAGTGTTCATAAAAATCATCGTTACAAAGATG 420  
DB 377 TGTACTCTGTATGGACTGCTCCAGACACAGTGTTCATAAAAATCATCGTTACAAAGATG 436  
QY 421 CATACTTCTACTGGAGAGGGTCTGTGACTGTGGAGACACAGAGGCGATGAAAACATGGC 480  
DB 437 CATACTTCTACTGGAGAGGGTCTGTGACTGTGGAGACACAGAGGCGATGAAAACATGGC 496  
QY 481 CCTTTTGTGTAAATCATGAACCTGGAGAGAGGCTACTATATAAGAGAAATTCACGCTGT 540  
DB 497 CCTTTTGTGTAAATCATGAACCTGGAGAGAGGCTACTATATAAGAGAAATTCACGCTGT 556  
QY 541 CCCTTGAATGAAGAGGTAATTTGTCCAAAGCCAGGAAAATATTTCTTCAGTGATAAAATAT 600  
DB 557 CCCTTGAATGAAGAGGTAATTTGTCCAAAGCCAGGAAAATATTTCTTCAGTGATAAAATAT 616  
QY 601 GTCGTAGAAATGACTATATGGGAGAGGAAAAGAACTGCTCTGAACTCCAGATAAGG 660  
DB 617 GTCGTAGAAATGACTATATGGGAGAGGAAAAGAACTGCTCTGAACTCCAGATAAGG 676  
QY 661 KNR-----YVCVNDHHSYDH 675  
DB 677 GAGAAAATGAAAGATACTATTGTGCTCTTTCAATGATGAACACCAATTCATATGACCAC 736  
QY 676 GTCATATACAGCCTACAAAGAGCTCTTGACTGTGAGCTCCAGAGGCCAGTTGCATACC 735  
DB 737 GTCATATACAGCCTACAAAGAGCTCTTGACTGTGAGCTCGCAGAGGCCAGTTGCATACC 796  
QY 736 ACTGCCATTGACAAAGAGGCTGCTGGGCTGTTAAAGCGGAGCTTATGCTGCTGGCCAG 795  
DB 797 ACTGCCATTGACAAAGAGGCTGCTGGGCTGTTAAAGCGGAGCTTATGCTGCTGGCCAG 856  
QY 796 GAAGCAAGGAAGATATAAGAGTCAATCAGAAAATGTCTCTCAACATCCACTTCATGTA 855  
DB 857 GAAGCAAGGAAGATATAAGAGTCAATCAGAAAATGTCTCTCAACATCCACTTCATGTA 916  
QY 856 GAAGTATTACACTCAGAGATTATGGCTCATCAGAAATTTGCTTGGCTGCTGGTCTGCG 915  
DB 917 GAAGTATTACACTCAGAGATTATGGCTCATCAGAAATTTGCTTGGCTGCTGGTCTGCG 976  
QY 916 ATGAACAAAATTATGAGCTATTCAAGTGACTTTAGGCAGATCTTTGCCAAGCATGCCCT 975  
DB 977 ATGAACAAAATTATGAGCTATTCAAGTGACTTTAGGCAGATCTTTGCCAAGCATGCCCT 1036  
QY 976 AGAAGAAACCTGACTCGGAGAAATCCCTGCTCATAGCAGGTTAATGCTTTGGGATGCA 1035  
DB 1037 AGAAGAAACCTGACTCGGAGAAATCCCTGCTCATAGCAGGTTAATGCTTTGGGATGCA 1096  
QY 1036 AAGCTTTATAAGGTGCCGTAAGATCCCTCATGAATTCATCAGCAGTTTATTTATG 1095  
DB 1097 AAGCTTTATAAGGTGCCGTAAGATCCCTCATGAATTCATCAGCAGTTTATTTATG 1156  
QY 1096 GAGATGGAATACAAAACTCTTGTCTATGGAATTTGCAAGTATTATAAACAACGAG 1155  
DB 1157 GAGATGGAATACAAAACTCTTGTCTATGGAATTTGCAAGTATTATAAACAACGAG 1216  
QY 1156 AAAGAATATACAGTGATGATCATGACAGAACTATCTCTATAACTGCATTTTCAGTTTCAG 1215  
DB 1217 AAAGAATATACAGTGATGATCATGACAGAACTATCTCTATAACTGCATTTTCAGTTTCAG 1276  
QY 1216 ATGTTTACTGTTCTGCTCGACATCTTATTGAAGAGCAGAAATGTTATCTCTGTC 1275  
DB 1277 ATGTTTACTGTTCTGCTCGACATCTTATTGAAGAGCAGAAATGTTATCTCTGTC 1336  
QY 1276 ATTACTGAACCTCTGATAGAGTTTACTGAGTACTTGGACAGGAAACAATAAATTCAC 1335  
DB 1337 ATTACTGAACCTCTGATAGAGTTTACTGAGTACTTGGACAGGAAACAATAAATTCAC 1396

QY 1336 TTCCAGGGTTATAGCCAGGACAAAATTTGGGAAGAGTATATGACAGTAAATATGACCTAAAG 1395  
DB 1397 TTCCAGGGTTATAGCCAGGACAAAATTTGGGAAGAGTATATGACAGTAAATATGACCTAAAG 1456  
QY 1396 TATATCCTGATCAGCAAAACCCACAATATGACAGAAAGATTAAAGATTGCAATTCCTTGAA 1455  
DB 1457 TATATCCTGATCAGCAAAACCCACAATATGACAGAAAGATTAAAGATTGCAATTCCTTGAA 1516  
QY 1456 GGTTTTTCGATCTTTTGAAGATTCTTACCTGTATGACGGGAATGGAAGAAATCCGAAGA 1515  
DB 1517 GGTTTTTCGATCTTTTGAAGATTCTTACCTGTATGACGGGAATGGAAGAAATCCGAAGA 1576  
QY 1516 CAGTTTGGCAACACATTGAAGTGGATCTGATTTGGAGGCTGCCATTCGTATACAGATG 1575  
DB 1577 CAGTTTGGCAACACATTGAAGTGGATCTGATTTGGAGGCTGCCATTCGTATACAGATG 1636  
QY 1576 CAATTGAAGAATATTTTACTCATGTTTCCAAAGAGTGTGTGCTTGTGATCAAGAATCTTTA 1635  
DB 1637 CAATTGAAGAATATTTTACTCATGTTTCCAAAGAGTGTGTGCTTGTGATGAAGAATCTTTA 1696  
QY 1636 CTGTGGCTTATAAAGAATGTCACAAAGCTGTGATGAGTGCGAGTACCAGTTTTCATATCT 1695  
DB 1697 CTGTGGCTTATAAAGAATGTCACAAAGCTGTGATGAGTGCGAGTACCAGTTTTCATATCT 1756  
QY 1696 AGTAGAAGACAGTAGTACAACTCGTGTGGACATAGTTTGGAAACAAAGTCCCTACAGATG 1755  
DB 1757 AGTAGAAGACAGTAGTACAACTCGTGTGGACATAGTTTGGAAACAAAGTCCCTACAGATG 1816  
QY 1756 TCTGAGGATCTTGAAGCATACATCTGCCACTCTCTAGACCTTGTGCTTGTGCTTGTGCTTGT 1815  
DB 1817 TCTGAGGATCTTGAAGCATACATCTGCCACTCTCTAGACCTTGTGCTTGTGCTTGTGCTTGT 1876  
QY 1816 CGTTTAGCAGGCTGGGTGCTGTTTCAAGACTGCAATGAATTTGTGCTTGTGAGGACTTTT 1875  
DB 1877 CGTTTAGCAGGCTGGGTGCTGTTTCAAGACTGCAATGAATTTGTGCTTGTGAGGACTTTT 1936  
QY 1876 CAAGTAGAGTACTAGTGGAAATATCCTTTACGTTGTCTGCTGCTGCTGCTGCTGCTGCTGCT 1935  
DB 1937 CAAGTAGAGTACTAGTGGAAATATCCTTTACGTTGTCTGCTGCTGCTGCTGCTGCTGCTGCT 1996  
QY 1936 GCTGAGATGTGGCGAAGAAATGGAGTCTCTTATTAGCCAGGCTGTTTATTACCAAGAT 1995  
DB 1997 GCTGAGATGTGGCGAAGAAATGGAGTCTCTTATTAGCCAGGCTGTTTATTACCAAGAT 2056  
QY 1996 GTTAAGTGCAGAGAAAGAAATGTATGATAAGAGATATCATGCTTCAGATTGGTGCATCT 2055  
DB 2057 GTTAAGTGCAGAGAAAGAAATGTATGATAAGAGATATCATGCTTCAGATTGGTGCATCT 2116  
QY 2056 TTAATGGATCCCAATAAGTTCTTGTACTGCTTACTGCTCAGAGGTATGAACCTTGGCGAGCT 2115  
DB 2117 TTAATGGATCCCAATAAGTTCTTGTACTTGTGCTGCTCAGAGGTATGAACCTTGGCGAGCT 2176  
QY 2116 TTTAAACAAGACCATATCTACAAAAGACGAGATTGATTAACAATATAATACACTATA 2175  
DB 2177 TTTAAACAAGACCATATCTACAAAAGACGAGATTGATTAACAATATAATACACTATA 2236  
QY 2176 GAAGAAATGCTTCAGGTCCTCATCTATATTGTGGGTGAGCGTTATGTACCTGGAGTGGGA 2235  
DB 2237 GAAGAAATGCTTCAGGTCCTCATCTATATTGTGGGTGAGCGTTATGTACCTGGAGTGGGA 2296  
QY 2236 AATGTGACCAAGAGAGGTCACAATGAGAAATCATTCACTTGCTTGTGCTTGTGCTTGTGCTTGT 2295  
DB 2297 AATGTGACCAAGAGAGGTCACAATGAGAAATCATTCACTTGCTTGTGCTTGTGCTTGTGCTTGT 2356  
QY 2296 ATGCCACAGTCCCATTTGCCAAAATTTTACCTGAGAAATGAANAATGAACCTGGCTTA 2355  
DB 2357 ATGCCACAGTCCCATTTGCCAAAATTTTACCTGAGAAATGAANAATGAACCTGGCTTA 2416  
QY 2356 GAGAAATGCTTCAAAACAAGTGGCCACATTTTAAAGAAACAGGTGTATCAGGCCATGGAGTT 2415  
DB 2417 GAGAAATGCTTCAAAACAAGTGGCCACATTTTAAAGAAACAGGTGTATCAGGCCATGGAGTT 2476  
QY 2416 TATGAACATAAAGATGAATCACTGAAGACTTCAATATGACTTTTATCATTTACTCCAAA 2475

[illegible]

Db 901 GCAGGACAAATTGGGAAGAGTATATGCAGTAATATGTGACCTAAAGTATATCCTGATCA 960  
QY 1409 GCAAAACCCACAATATGACAGAAAGATTAAGATGAGTTCCTTGAGAGTTTCGATCTT 1468  
Db 961 GCAAAACCCACAATATGACAGAAAGATTAAGATGAGTTCCTTGAGAGTTTCGATCTT 1020  
QY 1469 TTTTGAAGATTCTTACCTGTATGACAGGAATGGAAGAAATCCGAAGACAGGTTGGGCAAC 1528  
Db 1021 TTTTGAAGATTCTTACCTGTATGACAGGAATGGAAGAAATCCGAAGACAGGTTGGGCAAC 1080  
QY 1529 ACATTGAAGTGGATTCGATGTTGGAGGCTGCCATTGCTATACAGATGCAATTAAGAATA 1588  
Db 1081 ACATTGAAGTGGATTCGATGTTGGAGGCTGCCATTGCTATACAGATGCAATTAAGAATA 1140  
QY 1589 TTTTACTCATGTTCCAAAGAGTGTGCTTGTGATGAAGAACTTCTTACTTGTGGCTTATA 1648  
Db 1141 TTTTACTCATGTTCCAAAGAGTGTGCTTGTGATGAAGAACTTCTTACTTGTGGCTTATA 1200  
QY 1649 AAGAATGTCACAAAGCTGTGATGAGTGCAGTACCAGTTTTCATATCTAGTAGCAAGACAG 1708  
Db 1201 AAGAATGTCACAAAGCTGTGATGAGTGCAGTACCAGTTTTCATATCTAGTAGCAAGACAG 1260  
QY 1709 TAGTCAATGCTGTGACATAGTTTGGAAACAAAGTCCCTACAGAGTATCTGAGGATCTTG 1768  
Db 1261 TAGTCAATGCTGTGACATAGTTTGGAAACAAAGTCCCTACAGAGTATCTGAGGATCTTG 1320  
QY 1769 TAAGCATACATCTGCCACCTCTTAGACCCCTTCTGCTTCTATGATGAGTTCAGGATGAG 1828  
Db 1321 TAAGCATACATCTGCCACCTCTTAGACCCCTTCTGCTTCTATGATGAGTTCAGGATGAG 1380  
QY 1829 TGGGTGCTGTTTCAAGACTGCATGAATTTGCTGCTTTTGGAGGACTTTCAAGTAGAGTAC 1888  
Db 1381 TGGGTGCTGTTTCAAGACTGCATGAATTTGCTGCTTTTGGAGGACTTTCAAGTAGAGTAC 1440  
QY 1889 TAGTGAATATCCTTTACGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1948  
Db 1441 TAGTGAATATCCTTTACGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1500  
QY 1949 GAAGAAATGGAGTCTCTCTTATAGCCAGGTGTTTATACCAAGATGTTAAAGTGCAGAG 2008  
Db 1501 GAAGAAATGGAGTCTCTCTTATAGCCAGGTGTTTATACCAAGATGTTAAAGTGCAGAG 1560  
QY 2009 AAGAAATGATATAAGATATCATCATGCTTCAGATGCTGCTGCTGCTGCTGCTGCTGCTGCT 2068  
Db 1561 AAGAAATGATATAAGATATCATCATGCTTCAGATGCTGCTGCTGCTGCTGCTGCTGCTGCT 1620  
QY 2069 ATAAGTCTTGTGTTACTGTTACTTCAGAGGTATGAACCTGCGAGGCTTTTAAACAAGACCA 2128  
Db 1621 ATAAGTCTTGTGTTACTGTTACTTCAGAGGTATGAACCTGCGAGGCTTTTAAACAAGACCA 1680  
QY 2129 TATCTACAAAGACCGAGTATGATTAAACAATATATACATCACTAAATAGAAGAAATGCTTC 2188  
Db 1681 TATCTACAAAGACCGAGTATGATTAAACAATATATACATCACTAAATAGAAGAAATGCTTC 1740  
QY 2189 AGTCTCTCATCTATATTGTGGGTGAGGTTATGCTGAGTGGGAAATGTGACCAAG 2248  
Db 1741 AGTCTCTCATCTATATTGTGGGTGAGGTTATGCTGAGTGGGAAATGTGACCAAG 1800  
QY 2249 AAGAGTTCACATGAGAGAAATCATTCACCTTGTCTTTCATGTAACCCATGCCACACAGTG 2308  
Db 1801 AAGAGTTCACATGAGAGAAATCATTCACCTTGTCTTTCATGTAACCCATGCCACACAGTG 1860  
QY 2309 CCATTGCCAAAATTTACCTGGAATGAAATTAATGAACTGGCTTGAAGAAATGTCTATAA 2368  
Db 1861 CCATTGCCAAAATTTACCTGGAATGAAATTAATGAACTGGCTTGAAGAAATGTCTATAA 1920  
QY 2369 ACAAGTGGCCATTTAAGAACCCAGGTGTATCAGGCCATGGATTTATGAACATAAG 2428  
Db 1921 ACAAGTGGCCATTTAAGAACCCAGGTGTATCAGGCCATGGATTTATGAACATAAG 1980  
QY 2429 ATGAATCACTGAAGACTTCAATATGATGATCTTTTATCTACTTATCTACTTCAAAACCCAGCATAGCA 2488  
Db 1981 ATGAATCACTGAAGACTTCAATATGATGATCTTTTATCTACTTATCTACTTCAAAACCCAGCATAGCA 2040

QY 2489 AGGCTGRACATATGCAGAGAAAGGAGCAAAACAAAGAAACAAAGATGAAGCATTTGCCGC 2548  
Db 2041 AGGCTGAACATATGCAGAGAAAGGAGCAAAACAAAGAAACAAAGATGAAGCATTTGCCGC 2100  
QY 2549 CACCACCACTCCCTGGAATTTCTGCCCTGCTTTTCAGCAAAAGTGAATTAACCTTCTCAACTG 2608  
Db 2101 CACCACCACTCCCTGGAATTTCTGCCCTGCTTTTCAGCAAAAGTGAATTAACCTTCTCAACTG 2160  
QY 2609 ATATCATGATGATACATCTCAGGACCGCTATTTCAGCGGCAATAGACACAGATTTCTAACT 2668  
Db 2161 ATATCATGATGATACATCTCAGGACCGCTATTTCAGCGGCAATAGACACAGATTTCTAACT 2220  
QY 2669 TGTGGACCGAAGGATGCTCCAAATGGCTTTTCATATTTCTGCCATTTGGTTTACTAGAG 2728  
Db 2221 TGTGGACCGAAGGATGCTCCAAATGGCTTTTCATATTTCTGCCATTTGGTTTACTAGAG 2280  
QY 2729 AGAAGCAACAGCTTCAAAAGCTCCTGAAGAAAGTAAACATTTTACATTTTATCATTAAG 2788  
Db 2281 AGAAGCAACAGCTTCAAAAGCTCCTGAAGAAAGTAAACATTTTACATTTTATCATTAAG 2340  
QY 2789 CTTCAGATTTGGGAAGTTTCAGCCATGAATATACAAATGCTTTTGGAAAACTCAAGGAA 2848  
Db 2341 CTTCAGATTTGGGAAGTTTCAGCCATGAATATACAAATGCTTTTGGAAAACTCAAGGAA 2400  
QY 2849 TTCCCCAGTTAGAGGCCAGAGGACATGATAACGTGGATCTTTCAGATGTTTGACACAG 2908  
Db 2401 TTCCCCAGTTAGAGGCCAGAGGACATGATAACGTGGATCTTTCAGATGTTTGACACAG 2460  
QY 2909 TGAAGCGATTAAAGAGAAAAATCTGTTTAAATTTAGAACACACATCAGGATCGGAATCTA 2968  
Db 2461 TGAAGCGTTAAAGAGAAAAATCTGTTTAAATTTAGAACACACATCAGGATCGGAATCTA 2520  
QY 2969 TTAAGATGATGAGATTTACTCATGATAAG 2998  
Db 2521 TTAAGATGATGAGATTTACTCATGATAAG 2550

## RESULT 9

AK027803

LOCUS

DEFINITION Homo sapiens cDNA FLJ14897 fis, clone PLACE1004743, weakly similar to PROBABLE N-END-RECOGNIZING PROTEIN.

ACCESSION

AK027803

VERSION

AK027803.1

KEYWORDS

oligo capping; fis (full insert sequence).

SOURCE

Homo sapiens

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1

AUTHORS

Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,

Nishikawa,T., Nagai,K., Sugano,S., Aotsuka,S., Yoshikawa,Y.,

Matsunawa,H., Ishii,S., Kawai,Y., Saito,K., Yamamoto,J.,

Wakamatsu,A., Nakamura,Y., Nagahari,K., Masuho,Y. and Sasaki,N.

TITLE

NEDO human cDNA sequencing project

JOURNAL

Unpublished

REFERENCE

2

AUTHORS

Isogai,T. and Otsuki,T.

TITLE

Direct Submission

JOURNAL

Submitted (10-MAY-2001)

COMMENT

Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan

(E-mail: genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)

NEDO human cDNA sequencing project supported by Ministry of

Economy, Trade and Industry of Japan; cDNA full insert sequencing;

Research Association for Biotechnology; cDNA library construction,

5'- &amp; 3'-end one pass sequencing and clone selection; Helix

Research Institute (supported by Japan Key Technology Center etc.)

and Department of Virology, Institute of Medical Science,

University of Tokyo.

Location/Qualifiers

1..2550

/organism="Homo sapiens"

```

/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="PLACE1004743"
/tissue_type="placenta"
/clone_lib="PLACE1"
/note="cloning vector: pME18FL3"
117..>2550
/note="unnamed protein product"
/codon_start=1
/protein_id="BAB55380.1"
/db_xref="GI:14042752"
/translation="MTWEEGRKVPPELQIREK
ALDCLAEALQHTTATIDEEGRKAVKAGAACQ
ETMAHQKALRLGSWANKIMSYSDRQIFCOAQ
KGARKLHLIFSSFFEMSEYKKLFAMEYKYY
FVPTLARHLETEQNVSVTETLLELVPEYLD
KYLTKPTIETRLMQFLEGRPSFLVTKTCM
IOMALKNLLMFQEWCAQDELLLVYAKCHKA
KSTRVSEDVLSHLPSRTLAGUHLVRLSGAV
LVQVAVQAMRRNGLSLSQSVYQDVKREEL
LQRYELAEAFNTSTKDQDLKQYNTLEML
RETIHLICIDEPHUSATKRNENNETFGLN
DLFDNNVYHYKSTQHSKAEHQKRRQENHL
IMMYLTVFERADTDSNLWTGMLKQAFHLL
KASRLGSAMNITMLLEKLAGLPQLEGQDMITI
SESLKNDTEHDK"

```

Query Match	46.8%;	Score 2434.4;	DB 9;	Length 2550;
Best Local Similarity	97.6%;	Pred. No. 0;		

QY	494	ATCATGAACCTGGAAGACGAGGTACTATATAAAGAGAAATTCACGCTGCTCCGTTGAATGAAG	553
Db	1	ATCATGAACCTGGAAGACGAGGTACTATATAAAGAGAAATTCACGCTGCTCCGTTGAATGAAG	60
QY	554	AGGTAATTGTCACGAGCAAAATATTTCTTCAGTGATAAAATATGTCGTAGAAATGA	613
Db	61	AGGTAATTGTCACGAGCAAAATATTTCTTCAGTGATAAAATATGTCGTAGAAATGA	120
QY	614	CTATATGGGAGAGGAAAAAGAACTGCTCTCGAACCTCCAGATCAAGKNR	663
Db	121	CTATATGGGAGAGGAAAAAGAACTGCTCTCGAACCTCCAGATCAAGGAGAAAAATGAAA	180
QY	664	-----YYCVNDHSHYDHGTCATATACAGCC	688
Db	181	GATACTATTGTGCTCTTTTCAATGATGAACACCATTCATATGACGAGCTATATACAGCC	240
QY	689	TACAAAGAGCTCTTGACTGTGAGCTCGCAGAGGCCAGTTGCATACCACATGCCATTGACA	748
Db	241	TACAAAGAGCTCTTGACTGTGAGCTCGCAGAGGCCAGTTGCATACCACATGCCATTGACA	300
QY	749	AAGAGGGTCCTCGGGCTGTTTAAAGCGGGAGCTTATGCTGCTGCGCAGGAAGCAAAAGGAAG	808
Db	301	AAGAGGGTCCTCGGGCTGTTTAAAGCGGGAGCTTATGCTGCTGCGCAGGAAGCAAAAGGAAG	360
QY	809	ATATAAGAGTCATTCAGAAAATGCTCTCAACATCCACTTCATGTAGAGTATTACACT	868
Db	361	ATATAAGAGTCATTCAGAAAATGCTCTCAACATCCACTTCATGTAGAGTATTACACT	420
QY	869	CAGAGATTANGGCTCATCAGAAAATTTCTTTGGGCTTGTCTTCGTGATGAACAAAAATTA	928
Db	421	CAGAGATTANGGCTCATCAGAAAATTTCTTTGGGCTTGTCTTCGTGATGAACAAAAATTA	480
QY	929	TGAGCTATTCAAGTGCATTTTAGGCAGATCTTTTGCCAAAGCATGCTTTAGAGAAGAACTG	988
Db	481	TGAGCTATTCAAGTGCATTTTAGGCAGATCTTTTGCCAAAGCATGCTTTAGAGAAGAACTG	540
QY	989	ACTCGGAGAAATCCCTGTCATTAAGCAGGTTAATGCTTTGGGATGCAAAAGCTTTTATAAG	1048
Db	541	ACTCGGAGAAATCCCTGTCATTAAGCAGGTTAATGCTTTGGGATGCAAAAGCTTTTATAAG	600

Qy	1049	GTGCCGTAAGATCCCTTCATGAATTTGATCTTCAGCAGCTTTT	1108
Db	601	GTGCCGTAAGATCCCTTCATGAATTTGATCTTCAGCAGCTTTT	660
Qy	1109	AAAAACTCTTTTGTCTATGGAATTTGTGAAGTATTATAAACAACTGCACAAAGAATATATATCA	1168
Db	661	AAAAACTCTTTTGTCTATGGAATTTGTGAAGTATTATAAACAACTGCACAAAGAATATATATCA	720
Qy	1169	GTGATGATCATGACAGAAGTATCTCTATAACTGCACATTTCAGTTTCAGATGTTTACTGCTTC	1228
Db	721	GTGATGATCATGACAGAAGTATCTCTATAACTGCACATTTCAGTTTCAGATGTTTACTGCTTC	780
Qy	1229	CTACTCTGGCTCGACATCTTATTGAAGACAGAAATGTTATCTCTGTCATTACTCGAAACTC	1288
Db	781	CTACTCTGGCTCGACATCTTATTGAAGACAGAAATGTTATCTCTGTCATTACTCGAAACTC	840
Qy	1289	TGCTAGAAGTTTACCTGACTACTGTGCAGAGGAACTAAATTCACACTCCACGGTTTATA	1348
Db	841	TGCTAGAAGTTTACCTGACTACTGTGCAGAGGAACTAAATTCACACTCCACGGTTTATA	900
Qy	1349	GCCAGGACAAATTTGGGAAGAGTATATGCAGTAAATATGTGACCTAAAGTATATCTCTGATCA	1408
Db	901	GCCAGGACAAATTTGGGAAGAGTATATGCAGTAAATATGTGACCTAAAGTATATCTCTGATCA	960
Qy	1409	GCAACCCCAATATGGACAGAAAGATTAGAATGCGAGTTCCTTTGAAGGTTTTCGATCTTT	1468
Db	961	GCAACCCCAATATGGACAGAAAGATTAGAATGCGAGTTCCTTTGAAGGTTTTCGATCTTT	1020
Qy	1469	TTTTTGAAGATTCCTACCTGTTATGCAGGGAATGGAAGAAATCCGNAGACAGGTTGGGCAAC	1528
Db	1021	TTTTTGAAGATTCCTACCTGTTATGCAGGGAATGGAAGAAATCCGNAGACAGGTTGGGCAAC	1080
Qy	1529	ACATTGAAGTGGATCTCTGATTGGGAGGCTGCCAATTTGCTATTACAGATGCAATTCGAAGAATA	1588
Db	1081	ACATTGAAGTGGATCTCTGATTGGGAGGCTGCCAATTTGCTATTACAGATGCAATTCGAAGAATA	1140
Qy	1589	TTTTTACTCATGTTTCCAAAGAGTGGTGCTTTGATGAAGAACTCTTACTTTGGCTTTATA	1648
Db	1141	TTTTTACTCATGTTTCCAAAGAGTGGTGCTTTGATGAAGAACTCTTACTTTGGCTTTATA	1200
Qy	1649	AAGAAATGCACAAGCTGTGATGAGTGCGAGTACCAGCTTTCATATCTAGTAGCAAGACAG	1708
Db	1201	AAGAAATGCACAAGCTGTGATGAGTGCGAGTACCAGCTTTCATATCTAGTAGCAAGACAG	1260
Qy	1709	TAGTACAATCGTGTGGACATAGTTTGGAAACAAGTCCCTACAGAGTATCTCAGGATCTTG	1768
Db	1261	TAGTACAATCGTGTGGACATAGTTTGGAAACAAGTCCCTACAGAGTATCTCAGGATCTTG	1320
Qy	1769	TAGCATATCATCTGCCACTCTCTAGGACCCCTGCTGGTCTTCATGTAAGTAAAGCAGGC	1828
Db	1321	TAGCATATCATCTGCCACTCTCTAGGACCCCTGCTGGTCTTCATGTAAGTAAAGCAGGC	1380
Qy	1829	TGGGTGCTGTTTCAAGACTGCATCAATTTCTGTCTTTTGGAGCTTTCAAGTAGAGCTAC	1888
Db	1381	TGGGTGCTGTTTCAAGACTGCATCAATTTCTGTCTTTTGGAGCTTTCAAGTAGAGCTAC	1440
Qy	1889	TAGTGGAAATACCTTTACGTTTGTCTGGTGTGGTGGCCAGGTTGTTGCTGAGATGTGGC	1948
Db	1441	TAGTGGAAATACCTTTACGTTTGTCTGGTGTGGTGGCCAGGTTGTTGCTGAGATGTGGC	1500
Qy	1949	GAAGAAATGACCTCTCTTTATTAGCCAGGTTGTTTATTACCAAGATGTTTAAAGTCAGAG	2008
Db	1501	GAAGAAATGACCTCTCTTTATTAGCCAGGTTGTTTATTACCAAGATGTTTAAAGTCAGAG	1560
Qy	2009	AAGAAATGATGATAAAGATATCATCATGCTTCAGATTTGGTGCACTTTTAAATGGATCCCA	2068
Db	1561	AAGAAATGATGATAAAGATATCATCATGCTTCAGATTTGGTGCACTTTTAAATGGATCCCA	1620
Qy	2069	ATAAGTCTCTTTACTGGTACTCTCAGAGGTTGAACCTTGCAGGCTTTTAAACAAGACCA	2128
Db	1621	ATAAGTCTCTTTACTGGTACTCTCAGAGGTTGAACCTTGCAGGCTTTTAAACAAGACCA	1680
Qy	2129	TATCTACAAAGACCCAGGATTTGATTAAACAATATAATACACTAAATAGAAGAATGCTTC	2188

Db	1681	TATCTACAAAGACCGAGGATTTGATTAAACAATATAACACATTAATAGAGAAATGCTTC	1740
QY	2189	AGGTCTCATCTATATTTGGGTGACCGTATGTACTCTGAGTGGGAAATGTGACCAAAAG	2248
Db	1741	AGGTCTCATCTATATTTGGGTGACCGTATGTACTCTGAGTGGGAAATGTGACCAAAAG	1800
QY	2249	AAGAGTGCACATGAGAGAAATCATCTGCTTTGGCATTTGAACCCATGCCACACAGTG	2308
Db	1801	AAGAGTGCACATGAGAGAAATCATCTGCTTTGGCATTTGAACCCATGCCACACAGTG	1860
QY	2309	CCATTGCCAAAATTTACCTGAGAAATGAAAATGAATGAAATGCTGCTTTAGAGAATGTCATAA	2368
Db	1861	CCATTGCCAAAATTTACCTGAGAAATGAAAATGAATGAAATGCTGCTTTAGAGAATGTCATAA	1920
QY	2369	ACAAAGTGCCACATTTAAGAACACAGGTGTATCAGGCCATGAGGTTTATGAACATAAAAG	2428
Db	1921	ACAAAGTGCCACATTTAAGAACACAGGTGTATCAGGCCATGAGGTTTATGAACATAAAAG	1980
QY	2429	ATGAATCACTGAAGACATTTCAATATGCTATTTATCATTTACTCCAAACCCAGCATAGCA	2488
Db	1981	ATGAATCACTGAAGACATTTCAATATGCTATTTATCATTTACTCCAAACCCAGCATAGCA	2040
QY	2489	AGGCTGAACATATGCAAGAAAGAGAGAAACAAAGAAATGAAGCATTCGCCG	2548
Db	2041	AGGCTGAACATATGCAAGAAAGAGAGAAACAAAGAAATGAAGCATTCGCCG	2100
QY	2549	CACCACCACTCCTGAATTTCTGCCCTGCTTTTCAGCAAAAGTGATTAACCTTCTCAACTGTG	2608
Db	2101	CACCACCACTCCTGAATTTCTGCCCTGCTTTTCAGCAAAAGTGATTAACCTTCTCAACTGTG	2160
QY	2609	ATATCATGATGTACATTTCTCAGACCGTATTTGAGCGGCAATAGACAGATTTCTAACT	2668
Db	2161	ATATCATGATGTACATTTCTCAGACCGTATTTGAGCGGCAATAGACAGATTTCTAACT	2220
QY	2669	TGTGGACCAAGGATGCTCCAAATGGCTTTTCATATTTCTGCAATTTGGGTTTACTAGAAG	2728
Db	2221	TGTGGACCAAGGATGCTCCAAATGGCTTTTCATATTTCTGCAATTTGGGTTTACTAGAAG	2280
QY	2729	AGAAGCAACAGCTTCAAAAAGCTCCTGGAAGAAGTAACATTTGACTTTTATCATAAAG	2788
Db	2281	AGAAGCAACAGCTTCAAAAAGCTCCTGGAAGAAGTAACATTTGACTTTTATCATAAAG	2340
QY	2789	CTTCAAGATTGGGAAGTTTCAAGCATGAATATACAAATGCTTTTGGAAAACTCAAGAGAA	2848
Db	2341	CTTCAAGATTGGGAAGTTTCAAGCATGAATATACAAATGCTTTTGGAAAACTCAAGAGAA	2400
QY	2849	TTTCCCAAGTTAGAGGCCAGAGGACATGATAAGCTGGGATCTTCAAGATGTTTGACACAG	2908
Db	2401	TTTCCCAAGTTAGAGGCCAGAGGACATGATAAGCTGGGATCTTCAAGATGTTTGACACAG	2460
QY	2909	TGAAGCGATTAAAGAGAAAATCTGTTTAATTTAGTGAACACCATCAGGATCGGAATCTA	2968
Db	2461	TGAAGCGTTAAAGAGAAAATCTGTTTAATTTAGTGAACACCATCAGGATCGGAATCTA	2520
QY	2969	TTAAGAAATGATGAGATTACTCATGATAAAG	2998
Db	2521	TTAAGAAATGATGAGATTACTCATGATAAAG	2550
RESULT 10			
AY061884			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
REFERENCE			
AUTHORS			

Novel ubiquitin ligase E3 alpha-II		Unpublished	
2 (bases 1 to 5268)		Han, H.Q. and Kwak, K.	
Direct Submission		Submitted (05-NOV-2001)	
Research, Amgen Inc., One Amgen Center Drive, Thousand Oaks, CA 91320, USA		Location/Qualifiers	
1. 5268		/organism="Homo sapiens"	
/mol_type="mrna"		/db_xref="taxon:9606"	
1. 5268		/codon_start=1	
/product="ubiquitin ligase E3 alpha-II"		/protein_id="AAL32101.1"	
/db_xref="GI:27434480"		/translation="MASELEPEVQAIIDRSLLSECSAEIAGKWLQATDLITREYVOHLAH	
YVPIKTCRGNPFPPQKEDMLAQHVLLGPMEMWLCGEDPAFGFPKLEQANKPSHLGGRV		FKYGEPTYSRCDCAVDPTCLCMCEFLGSHRDHRYRMTTSGGGGDCDGGTEAWKEG	
PYCKHELNTSEIEEDVPLVHSEDVIATYNTFAITFRYAVILTWKESELPADL		EMVEKSDTYCYMLFNDEVTYEOVYITLQKAVNTOKEAIGFATTVDRGRSVRYGD	
FOYCEQAKSVIVRNTSRQTKPLKVQVMHSSIVAHQNFGLKLLSWGLSIGYSDGLRRI		LCQVGLEGPDGNSSLVDRLMSDSLKWKGARSYYHQLPMSLLMOLKYKLLFAVRP	
AKNYQQIQRDMEDDHERAVSVTSLVSQVFFAPTAPTLARMLITEENMLIITKTFMDHLR		ELDKMQGMDPIRQVGHHEMEPEWAAFTLQMKLTHVISMMDQDCASDEKVIETAY	
KCLAVLMQCHGGYTDEQPIITLISGHVSVEITRYCVSOEKVSIHLPSVRLLAGLHVL		LKSSEVAYKPELLPLSELSPMILEHPLRCLVLCQVHAGMWRNNGFSLVNOIYVYH	
NVKCRREMFDKDVVMTQGVSMDDPHFLMIMLSRFLYQIFETPDGKRRFSSEITHK		SLDPEDNKETGMSVTEADAVAFKKPLGTGRGMVELKPEKAFNLYFYHFSRAEQLK	
EQAQRKLQKQNRREDTALPPVLPFCPLFASLVNLIQSDQNLQIMGTILQAWAHNGY		AWESMRLQVHLHIGALOEKOLENVTEHVYFTTOKISKPGAPKNSPILAM	
LETLOAPYLEVHKDMIRWILTKFNKVMRESSPTSPVAETGTIMEESSRDKDAE		RKKAETARLRREKIMQNMQRHFDENKELPQOTLELDASVSLDHSVPASDMT	
RTALGPAQTQPEQRQFVTCILCOEQEVESRAMVLAAPFQSTVLSNRSKFIQD		PEKYDPLFMHPDLSCGTHHTSCGHIMHAHQWRYFDSQAKERQRRQLRLHTSYDVE	
NGFGLPCLECLNTVTPLLPPRINFNNRLNFSQDQNLQIMGTILQAWAHNGY		EESTPNNAATKNSENVDELQPEGRFDPFRPKIPYSESIKEMLTFTGTATYKVLKVVH	
PNEEDPRVPIMCWGSCAYTIOSTERILSDRDKPLFGLPCRLDCLRLSTRFAAAHWT		VASVSVQGHCECKLFLASLPNDSSHEELPCILDIDMFHLLVGLVLFALQOCDFSGIS	
LGTGDLHLHLVTHAIIQILTLCTEENGMDQENPCEESAVSLYALKTHUQYTGSA		LKPEISGHWILSVRAGIMPFLLKCSALFFHLYLNGVSPDPDIQVGFSHFHEKLSL	
INNALICLFQENSEIMNSLIESWNSRSEKRYLEGERDAIRYPRESNKILNPEDYSSL		PNNASFPCKSGGDKSRAPTLCLVGLSLLCSQSCQCTQTELEDGKCAHTHTYSCGS	
GVGIFLVRRECQVLFELAGTKGCFYSPYLDYDGETDQGLRRGNPLHLCKRFRKIQK		LWQHVSYTEIIGHAQENQTLVGDWQHL"	
BASE COUNT		1618 a 1015 c 1193 g 1442 t	
ORIGIN			
Query Match		24.9%; Score 1294.2; DB 9; Length 5268;	
Best Local Similarity		55.6%; Pred. No. 7 8e-305;	
Matches 2904; Conservative		8; Mismatches 2157; Indels 150; Gaps 17;	
QY	85	TGGGATCAGCAAGTTGATTTTTTATCTGCTTTTTCATCTATTTTGGCACAATTTGGTGCCA	144
Db	82	TGGCTGCAAGCAACTGACCTCAGTAGAGAAGTGTACCGACATTTAGCCCACATGTACCC	141
QY	145	GAATTTACTTTGCTGAAATGGACCGACCTTGGAAAAGCAGGAGGAAAGTGTACAATG	204
Db	142	AAAATCTACTGACGGGTGCCAACCCCTTTTCCACAGAAAGAACATGCTGGCACAGCAT	201
QY	205	TCAATATTCACTCCACTGGAATGCTACTTATTTGGAGAAGATCCAGATATTTGCTTAGAG	264
Db	202	GTTTTGTGGACCAATGGAATGCTACTTTTGGTGAAGATCTCGCATTTGGATTTGCCA	261
QY	265	AAATTTGAAGCAC---AGTGGAGCATTTTCAGCTTTTGGGAGGGTGTTCACAAAGTGGAGAG	321
Db	262	AAACTTGACGACGACAAACACCTTCTCATCTTTGTGCTGCTGTTTTTAAAGTAGGAGAG	321
QY	322	ACAACCTATTTCTTGACGGGATTTGCAATTTGATCCCAACATGTGTACTCTGTATGCACTGC	381



Db 322 CCTACATATCTTGCAGAGACTGTGCAGTGTGATCCAACTTGTGTTTTGTGCATGCGTGC 381  
Qy 382 TTCCAGGACAGCTGTTCATAAAATCATCGTTACAGATGCATCTTCTACTTGGAGAGGG 441  
Db 382 TTTTGGGAATATTCACAGAGATCATCGATAGATGACAAACATCAGGAGGTGGAGGT 441  
Qy 442 TTCTGTGACTGTGAGACACAGAGCATGGAAACTGGCCCTTTTGTGTAATCATGAA 501  
Db 442 TTCTGTGACTGTGATCTGAAGCTTGAAGCTTGAAGAGGTCTTACTGTCAAAACATGAA 501  
Qy 502 CTTGGAAGAGAGAGTACT---ATAAAAGAGAAATTCACGCTCTCGCTTGAATGAAGAGTA 558  
Db 502 CTTAACACCTCTGAAATTTGAGAAAGAGAGATCCTCTGTGTTTATCATCAGAAGATGTG 561  
Qy 559 ATTGTCCAGCCAGGAAATATTTCCCTTCAGTGATAAAATATGTCGTAGAAATGACTATA 618  
Db 562 ATAGCAAGAACTTATACATTTTGTGCTATTACGTTTCGGTATGCGATGAGAAATATTAACC 621  
Qy 619 TGGGAAGAGGAAAGAACTGCCCTCCCTGAACTCCAGATAAGGKNRY----- 664  
Db 622 TGGGAAAGAGAAAGTGAATGCCAGCAGATTTAGAGATGGTAGAGAGAGTGACACCTTAC 681  
Qy 665 -----YCVNDHSHSDHGTGCATATACAGCCCTACAAA 693  
Db 682 TATTGCATGCTGTTTTAATGATGAGTTTCACACCTATGAACAAGTTATTTATCTCTCAG 741  
Qy 694 AGACCTCTGACTGTGACTCGCAGAGCCCGCTTGCATACACTGCCATTCACAAAGAG 753  
Db 742 AAAGCTTTAACTGTACACAAAAGAAAGCTATTGGTTTGTGAAGCTTACAGTAGATGGAT 801  
Qy 754 GGTGCTGGGTGTTAAAGCGGGAGCTTATGCTGCTGCCAGGAGCAAGAGATATA 813  
Db 802 GGGCGTAGTCTGTGATGAGATTTTCAGTATTGTGAGCAAGCAAAATCAGTAAT 861  
Qy 814 AAGAGTCATTACAGAAAATGCTCTCAACATCCACTTCATGTAGAGATTTACACTCAGAG 873  
Db 862 GTGAGAAATACCAGTAGACAGAC---AAAGCCACTCAAGTTCAGTTTGCATTCGCT 918  
Qy 874 ATTAAGCTCATCAGAAATTTGCTTTGGTCTTGGTCTCGATGAGCAAAATATAGAC 933  
Db 919 ATTGTCCACATCAGAAATTTGGTTTGAAGCTTTTGTGCTGGGAAATATTATTGA 978  
Qy 934 TATTCAAGTGACTTTAGGCAGATCTTTTGCCAAAGCATGCCCTTAGAGAGAACCCTGAC 993  
Db 979 TATTCAAGTGCCCTTCCCGGATTTTATGTCAGTTGGTTTACAGAAAGGCCAGATGGT 1038  
Qy 994 GAGATCCCTGTCTCATAAAGCAGGTAAATGCTTTGGGATGCAAGCTTTATAAGGTGCC 1053  
Db 1039 GAAACTCTTCTCTAGTGGACAGACTGATGCTTATGATTCCAAATTTAGGAAAGGTGCT 1098  
Qy 1054 CGTAAGATCCTTCATGAATGATCTTCAGCAGTTTTTTTTATGGAGATGGAATACAAAAA 1113  
Db 1099 AGGAGTCTATATCATCAGTTGTTTCATGAGCAGTCTGCTATGAGTTTGAATACAGAAA 1158  
Qy 1114 CTCTTTGCTATGGAATTTGGAAGTATTTATAAACAACCTGCAGAGAAGAATATCATAGT 1173  
Db 1159 CTAATTTGCTGTTGCAATTTGCAAAAAATTTACCAGCAGTTGCGACAGAGATTTTATG 1218  
Qy 1174 GATCATCAGAGAGTATCTCTATAACTGCACCTTTTCAGTTTCAGATGTTTACTGTTCC 1233  
Db 1219 GATCAGAGGAGCAGGTGCGGTGACTGCTCTATCTGTCCAGTTCTTCACCGCACCTACT 1278  
Qy 1234 CTGGCTCGACATCTTATTGAAGAGCAGAACTGTTATCTGTCTGCTTACTGAAACTCTGCTA 1293  
Db 1279 CTGGCTCGAATGCTCATCAGAGAAACCTTGATGAGCATTTATCAITTAAGACTTTTATG 1338  
Qy 1294 GAAGTTTT-----ACCTGAGTACTTGGACAGCAACATAAAATTCACCTCCAGGGTTAT 1347  
Db 1339 GATCATTTGAGACATCGAGATGCCCGGAGAGATTTTCAGTTTGAACGATACACTGCTTTA 1398  
Qy 1348 AGCCAGGACAAATTTGGAAGAGTATATGCAAGTAAATATGTGACCTTAAAGCTATATCCT 1407  
Db 1399 CAAGCCTTCAAAATTTAGGAGAGTACAGAGCCTTATTTTATAGATCTCAAGTATGTGTTA 1458

Qy 1408 AGCAAAACCACAAATATGAGCAGAAAGATTAAGAAATGCAGTTCCTTGTAGGCTTTTCGATCT 1467  
Db 1459 AGCAAAACCACAAATGATGGTCAGATGAGCTGAGCGAAGATTTCTTAGAAGGGTTTGTAGGCC 1518  
Qy 1468 TTTTGTGAAGATCTTACCTGTATGACAGGAATGGAAGAAATCCGAAGACAGGTGTGGCAA 1527  
Db 1519 TTTTGTGAATTAATAAATGATGACAGGAATGATGATGATGATGATGATGATGATGATGAT 1578  
Qy 1528 CACATTTGAAGTGGATCCTGATTTGGAGGCTGCCATTTGCTATACAGATGCAATTTGAAGAAT 1587  
Db 1579 CATATTTGAATGGAACACAGAGTGGGAAGCAGCTTACACTACAAATGAAATTAACACAT 1638  
Qy 1588 ATTTTACTCATGTTTCAAGAGTGGTGTGCTTGTGATGAGAAACTCTTACTTGTGGCTTAT 1647  
Db 1639 GTCAATTTCAATGATGAGGAGTGGTGTGCTTACAGATGAAAAGTGTTAATCGAAGCTTTAC 1698  
Qy 1648 AAGAAATGTCAAAAGCTGTGATGAGTGCAGTACCAGTTTCTATATCTAGTAGCAAGACA 1707  
Db 1699 AAGAAATGTCTCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1758  
Qy 1708 GT---AGTACAAATCGTGTGACATAGTTTTGGAAACAAAGTCTTACAGAGTATCTGAGGAT 1764  
Db 1759 ATCACAATAAGCATTTGTGACATTCAGTGGAAACTATCAGATACTGTGTTTCCCAAGAA 1818  
Qy 1765 CTTGTAAGCATACATCTGCCACTCTCTAGGACCCCTTGTGCTGCTTCTATGATGATGATGATGAT 1824  
Db 1819 AAGTTAGCATTTACCTCCAGTTTCTCGCTTACTTGCAGGTTTACATGATTTATTTAAGC 1878  
Qy 1825 AGGCTGGGTGCTTTCAAGACTGCATGAATTTGTGCTTTTGGAGGACTTTCAAGTAGAG 1884  
Db 1879 AAAAGTGAAGTGCATATAAATTTCCAGAGCTCCTACCTCTAAGTGAACCTTAGGCCACCC 1938  
Qy 1885 GTACTAGTGGAAATATCCTTTTACGTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1944  
Db 1939 ATGTTGATAGAAACACCTCTTAGATGCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1998  
Qy 1945 TGGGGAAGAAATGGAGTCTCTTATAGCAGGCTGTTTATTTACCAAGATGTTTAAGTGC 2004  
Db 1999 TGGAGAAGAAATGGGTCTCTCTAGTAAACACAGATTTATTTACTACCAATAATGTGAAATGC 2058  
Qy 2005 ACAGAAGAAATGATGATAAAGATATCATCATCTTCCAGATTTGGTGCATCTTTAATGAT 2064  
Db 2059 AGACGTGAGATGTTTGACAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2118  
Qy 2065 CCAATAAGTCTTGTGTTACTTGCAGAGTATGAGTATGAGTATGAGTATGAGTATGAGTATGAGTAT 2124  
Db 2119 CCAATCATTTCTGATGATCATGCTCAGCCGCTTTGAACTTTATCAGATTTTTCAGTACT 2178  
Qy 2125 ACCATATCTTACAAA-----GACCAGATTTTGAATTAACAATAT 2163  
Db 2179 CCAGACTATGGAAGAAAGATTTAGTTCTGAGATTTACCCATAAGGATGTTGTTTCAGCAGAAC 2238  
Qy 2164 AATACACTAATAGAGAAATGCTTCAGGCTCCTCATCTATATTTGTTGGGTGAGCGGTTATGTA 2223  
Db 2239 AATACTCTAATAGAGAAATGCTTATACCTCATTTAATGCTTGTGTTGAGAGAGATTTAGT 2298  
Qy 2224 CCGTGGGTGGAAATGTGACCAAGAGAGGTGCATATGAGAGAAATCTTCACTTCTGCT 2283  
Db 2299 CCGTGGGTGGAGGATTAATGCTACAGATGAAATCAAGCGAGAGATTTATCCATCAGTTG 2358  
Qy 2284 TGCATTTGAACCCATGCCACAGTGCCTTTGCCCCAAATTTTACTGAGATGAAATTAAT 2343  
Db 2359 AGTATCAAGCCTATGGCTCATAGTGAATTTGGTAAAGTCTTTACTGAGATGAGAACAG 2418  
Qy 2344 GAACTGGCTTAGAGAAATGTCATAAACAAGTGGCCACATTTTAAAGAACCCAGGTGTATCA 2403  
Db 2419 GAGACTGGCATGGAGAGTGTAACTGAAGCAGTTGCCCATTTTCAAGAAACCTGGATTAACA 2478  
Qy 2404 GGCATGGAGTTTGAACATAAAGATGAATCACTGAAAGACTTCAATATGTTACTTTTAT 2463  
Db 2479 GGACGAGGCGATGTGTAAGTGAACCCAGAAATGTCGAAGAGTTCACCTTGTATTTCTAT 2538

QY 2464 CATTACTCCAAAACCCAGCATAGCAAGGCTGAACATATGACAGAAAGAGAGAAACAA 2523  
D 2539 CACTTTTCAAGGCGACAGACAGTCCAGGCGAAGAGACCGCAACGGAATTTGAAAGACAA 2598  
QY 2524 GAAAACAAAGATGAAGCAATTCGCGCCACCAACCCCTCCTGAAATTCGCCCTGCTTTTCAGC 2583  
D 2599 AATAGAGAAGATACAGCACTCCCACTCCCGGTGTGCTTCCATTCGCCCTCTGTTGCA 2658  
QY 2584 AAGTGTATTAACCTTCTCACTGTGATATCATATGATGTACATTTCTCAGGACCGTATTTGAG 2643  
D 2659 AGCTCGTTAACAATTTTGCAGTCAGATGTCATGTTGTCATCATGGAACAATTCGCAA 2718  
QY 2644 CGGGCAATAGACACAGATTTCACTTGTGACCGAAGGATGCTCCAAATGGCTTTTCAT 2703  
D 2719 TGGGCTGTGGAACATTAATGATATGCTGTGTCAGATGCCATGCTGCAAGGGTGTATCAT 2778  
QY 2704 ATTCTGGCATTTGGGTTTACTAGAGAAGAACAGCACTTCAAAAAGCTCCTGAAG---AA 2760  
D 2779 TTAATTTGGCATGSCACTACAGAAGAAACACATTTAGAGATGTCAGGGAAGACAT 2838  
QY 2761 GAAGTAACATTTGACATTTTATCAATAGGCTTCAAGATTTGGAAGTTTCAAGCCATGAAT--- 2817  
D 2839 GTAGTAACATTTTACCTTCTCAGAAAGATATCAAAACCTGGTGAAGCGCCAAAATTTCT 2898  
QY 2818 -----ATACAAATGCTTTTGGAAAACCTCAAGGATTTCCCACTAGAGGCCAGAAG 2871  
D 2899 CCTAGCATACTAGTATGCTGGAACACACTACAAAATGCTCCCTACCTAGAGAGTCCACAA 2958  
QY 2872 GACATGATAACCTGGATCTCAGATGTTTGACACAGTGAAGCATTAAGAGAAAATCT 2931  
D 2959 GACATGATTCGGTGGATTTGAGACTTTTAAATGCTTTAAAGATGAGGAGAGATTCA 3018  
QY 2932 TGTTTAATTTAGCAACACATCAGGATCGGATCTTATTAAGAAATGATGAGATTACTCAT 2991  
D 3019 CCTACCAGTC-----CCGTGCGAGAGACAGAAAGCAACATAATGAAGAGAGTTCAAGG 3072  
QY 2992 GATAAAGAAAAGCAGACAGAAAGAAAGCTGAAGCTGTAGGCTACATCCGCCAGAAG 3051  
D 3073 GACAAAGCAAAAGCTGAGAGAGAGAAAGAGAGATTTGCCAGAGTGGCGAGAGAAAAG 3132  
QY 3052 ATCATGGCTCAGATGCTGCCCTTACAGAAAACCTTCATTGAAACTCATAACTCATGTAT 3111  
D 3133 ATCATGGCTCAGATGCTGAATGACGGCGCATTTTATGATGAAAACAAAGAACTCTTT 3192  
QY 3112 GACATPATCATGAAATGCTTGGGAAAGAAATTCATTTATGAGGAAAGAGAGACACCCCA 3171  
D 3193 CAGCAGACATTAAGAACTGGATGCTCAACCTCTGCTGTCTTGATCA-----TAGCCCT 3246  
QY 3172 GCAGTCAGTACTACTCTAGAATTTGCTTTGGGCTCTTAAGGGGGTCCATCTGTACTGAA 3231  
D 3247 GTGGCTTCAGATATGACACTTACAGCACTGGGCGCCGCACTCAGGTTCTCTGAACAA 3306  
QY 3232 AAGGAGGTGCTGACGTGTCCTTTGCCAAGAGAACAGGAGTGAATAAGAGAAATAAT 3291  
D 3307 AGACAATTCGTACATGTAATTTGTCGAAGGAGCAAGAGATTAAAGTGAAGACAGG 3366  
QY 3292 GCCATGGTATTATCGGCCCTGTGTCAGAAATCTACTGCCCTTAACCCAGCACAGGGGAAAA 3351  
D 3367 GCAATGGCTTGGCAGCAATTTGTTCAAGATCAACTGTATTTATCAAAAACAGAGATAA 3426  
QY 3352 CCCATAGAACTCTCAGGAGAAGCCCTAGACCACATTTTCATGATCCAGACTTGGCATAT 3411  
D 3427 TTTTATCAAGATCCAGAAAA---TATGATCCATTTATCATGCACTGATCTGTCTGT 3483  
QY 3412 GGAATTTATACAGGAGCTGTGTCATGTAATGTCAGGCACTGTGCTGGGCAAGATATTTT 3471  
D 3484 GGAACACACACTAGTAGCTGTGGGCAACATTTATGCAATGCTGTTGGCAAAAGGATTTT 3543  
QY 3472 GAAGCTGACACTGAGCTCTCAGCAGCGCAATTTTCATGTTGACCTTTT-----T 3519  
D 3544 GATTCGGTTCAAGCTTAAGAACAGCGAAGGCAACAGAGATTAGCTTTACATAGGAGCTAT 3603  
QY 3520 GACTTGGAAAGTGGAGAATATCTTTGCCCTCTTTGCAAAATCTCTGTGCAATACTGTGATC 3579

D 3604 GATAGAAAACGGGAATTCCTTTGCCCTTTGTGAAATGCTTGAGTAATAACTGTATT 3663  
QY 3580 CCCATTATTCCTTTGCAACCTTCAAAAGATAAACAGTGAAGATGCAGATGCTCTTGCTCAA 3639  
D 3664 CC-----TCTGCTGCTCTCCCAAGAAATATTTTAAACACAGAGTTAAATTTTTCAGAC 3717  
QY 3640 CTTTTGACCCCTGGCAGCGTGGATACAGACTGTTCTGGCCAGAAATATCAGGTTATATAATA 3699  
D 3718 CAACCAATCTGACTCAGTGGATTAGAACAAATATCTCAGCAAAATAAAAGCATTTACAGITT 3777  
QY 3700 AGACATGCTAAGAGAGAAAACCCAAATTCCTATTTCTTTTAAATCAAGGAATGGGAGATTCT 3759  
D 3778 CTTAGSAAAGAAAGAAAGTACTCTCTAATAATGCCCTCTACAAAAGAAATTCAGAAAATGGAT 3837  
QY 3760 ACTTTGGAGTTCCATCTCCTGAGTTTGGCGTTTGGCTGCTTCCATTAATAATATTAATAAT 3819  
D 3838 GAATTACAGCTCCCTGAAGGTTTCAGGCTGATTTTCTCTAGATCCCTTTATTTCTGAG 3897  
QY 3820 AGCATCAAGGAAATGGTTATTTCTCTTTGCCACAACAAATTTATAGAAATTTGGATTTGAAAGTG 3879  
D 3898 AGCATAAAGAAATGCTTAACGACATTTTGGAACTGCTACCTACAAGTGGGACTAAAGTT 3957  
QY 3880 CCACCTGATGAAGGGATCCTCGAGTCCCAATGCTGACCTGGAGCACCTGCGCTTTTCACT 3939  
D 3958 CATCCCAATGAAGAGGATCCTGCTGTTCCCAATAATGTTGGGTAGCTGGCGGTACACC 4017  
QY 3940 ATCCAGGCAATTTGAAATCTATTGCGGAGATGAAGAAAACCTCTGTTTGGAGCACTTCAA 3999  
D 4018 ATCCAAAGCATAGAAAAGAAATTTGAGTGAATGAAGATAAACCATTTGTTGGTCTTTACCT 4077  
QY 4000 AATAGGAGCATATGCTGTAAGACATTAATGAGATTTGAGATTTGCAGAGAGATTACC 4059  
D 4078 TGCAGACTGGATGACTGCTTAGTCTATTGAGAGATTTGCGCGACACACTGGACAGTG 4137  
QY 4060 TGTCTCAGGTCCTGATACAGAAACATCTGGTTCGTCTCTTATCAGTTGTTCTTCTCAAC 4119  
D 4138 GCATCAGTTTTCAGTGGTGAAGGACATTTTGTAAACTTTTGTGCATCACTGGTGCCTAAT 4197  
QY 4120 ATAAATCAGAAGATACACCATGCTCTGCTGCTATAGATCTGTTTCACTGTTTGGTGGGT 4179  
D 4198 GACAGCCATGAGGAGATTCATGATGATATAGATTTGACATGTTTATTTATTTGGTGGGC 4257  
QY 4180 GCTGTGTAGCATTTCCCATCTCTTATTTGGATGACCCCTGTTGATCTGCAGCCTTCTTCA 4239  
D 4258 TTGGTGTGCAATTTCTCTGCTGCTGAGTG-----TCAGATTTTTCAGGG 4302  
QY 4240 GTTAGTCTCTTATACACAGCTTTTATCTCTTCCATTTGATCAGCATGGCACAGATGTT 4299  
D 4303 ATCAGCCTTGGCACTGGAGACCTTTCACATTTTCCATCTGTTTACTATGGCACATCATA 4362  
QY 4300 CAGATCTACTTACAGTAGACACAGGCC---TACCCCTTGTCTCAGGTTCAAGAAAGACAGT 4356  
D 4363 CAGATCTTACTTACTCTATGTAAGAGAAATGGCATGGATCAAGAAAATCCCCCTTTGT 4422  
QY 4357 GAAGAGGCTCATTCGCGCATCTCTTTCTTTGCAAAAATTTCTCAATATACAAGTGGCTCC 4416  
D 4423 GAAGAAGAAATCAGCAGTTCTTGTCTGTATAAAACACTTACACAGTATACGGGAAGTCC 4482  
QY 4417 ATTGGGTGTGATATTCCTGGCTGGTATTTTGGGTCTCAGTGAAGAAATGGCATCACCCCT 4476  
D 4483 TTGAAGAAAATACCATCCGCTGGCATCTGTGGAGGAGTGTGCAGAGCTGGAATCATGCT 4542  
QY 4477 TATCTTCGCTGTGCTGATTTTTCCTCACTATTTTACTTGGGTAACTCCGCCCTGAGGAA 4536  
D 4543 TCCCTGAAGTCTGCTGCTTTATTTTTCATTTAAATGGAGTTCCTTCCCCACCCGAC 4602  
QY 4537 CTGATACCAATTTCTGAGAGAGAGATACAGTGCACCTCTGCTAGCTATCTATCTTTACT 4596  
D 4603 ATTCA---AGTTCCTGGAACAGCCATTTTGAACATTTATGTAGCTATCTTTCCCTACCA 4659  
QY 4597 ACAATTTTGTCTGCTCTTCCAGGAATTTGGGATATCTGTAAGGCCCTTCTCTCCAGAGG 4656

```
Db 4660 AACAACTCATTTGCCCTTTTCAAGAAAATAGTGAGATATGAATTCATCTGATGAAAGT 4719
QY 4657 CGGTGTGAGATCTCCCTTACTAAACTGTGTTGAAGCAAAAACACCCTGGTGACGTAC 4716
Db 4720 TGGTCCGTAACAGTGAAGTAAAGATATAGAAAGTCAAGAGATGCTATAAGATAT 4779
QY 4717 CCTAGAAAAGAAATAGTTGTAGAGCTTCTGATGACTATATAGCTGCTCCTGTAATCAA 4776
Db 4780 CCAAGAGAACTAAACAATTAATAACCTTCCAGAGGATATACAGAGCTCATTAATCAA 4839
QY 4777 GCTTCTCATTTGAGTGGCCACGGTCTGCAGATGATGAGGAAAGCATCTGTCCTCTGC 4836
Db 4840 GCATCAATTTCTGTCGCCGAAATCAGGTGGTGATAGAGCAGACCCCAACTCTGTGC 4899
QY 4837 CTTTCTGTGGGCTATATATGTTCTCAGAACATTTGCTGCCAGAAATTTGTGAACGGG 4896
Db 4900 CTTGTGCGGATCTCTGCTGCTGCCAGAGTTACTGCTGCCAGACTGAACCTGGAAGG 4959
QY 4897 GAAGAGTTGGAGCTGCAATTTTTCAGGCACTTCTGAGGAGCCGGAGTCTGCATTTTC 4956
Db 4960 GAGGATGTAGGAGCTGCAAGCTCACAGCTACCTCTGCTGGCTCTGGAGTGGCATCTTC 5019
QY 4957 CTAATAATCAGAGATCGGAGTGTCTGTTGAGGTAAAGCCAGAGGCTGTGCCTAT 5016
Db 5020 CTGAGAGTACGGGAATGTCAAGTGTCTATTTTAGCTGGCAAAACCAAGGCTGTTTTAT 5079
QY 5017 CCAGCTCTCTTACTTGGATGAATGTGAGAAACAGACCTTGGCTGAAGAGGGCAACCCC 5076
Db 5080 TCTCTCTCTTACTTGTGATGACTATGGGAGACCGACAGGACTCAGACGGGGAATCTCT 5139
QY 5077 CTTCAATTTATCTGTCGAGCGGTATCGGAAGCTCCATTTGCTGTGGCAACAACTGCATT 5136
Db 5140 TTACATTTATGCAAGAGCGATTCAGAGAGATTCAGAACTCTGGCACCAACACAGTGT 5199
QY 5137 ATAGAGAGATGTGTAGGAGCAAGAGACTAATCAGATGTTATTTGGATTCAACTGGCA 5195
Db 5200 ACAGAGGAATTTGGACATGCACAGGAAGCAATCAGACACTGTTTGGCATTTGACTGGCA 5258
```

## RESULT 11

```
AY061885
LOCUS Mus musculus 5265 bp mRNA linear ROD 30-DEC-2002
DEFINITION Mus musculus ubiquitin ligase E3 alpha-II mRNA, partial cds.
ACCESSION AY061885
VERSION AY061885.1 GI:27434481
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 5265)
AUTHORS Han,H.Q. and Kwak,K.
TITLE Novel ubiquitin ligase E3 alpha-II
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 5265)
AUTHORS Han,H.Q. and Kwak,K.
TITLE Direct Submission
JOURNAL Submitted (05-NOV-2001) Research, Amgen Inc., One Amgen Center
Drive, Thousand Oaks, CA 91320, USA
FEATURES
source
1..5265
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C3H/HeN"
/db_xref="taxon:10090"
/dev_stage="4-6 weeks old"
1..5265
/codon_start=1
/product="ubiquitin ligase E3 alpha-II"
/protein_id="AAL32102.1"
/db_xref="GI:27434482"
/translation="WASEMEPEVQADIRSLLECSAEIAGRWLQATDLNREVYQHLAH
CVPKIYCRGNPNFPQKEDTLAQHILGPMEWYICAEADPALGPFKLEQANKPSHLCGRV
```

CDS



QY 2753 CTGAGAGAAAGT---AACATTGACTTTTATCATAGAGCTTCAAGATTGGGAAGTTGAG 2809  
D 2828 TGGAGGCGACGTGCGAGACCTTACCTTCACACAGAGATTTCAAAGCGTGTGATGCAC 2887  
QY 2810 -----CCATGAATATACAAATGCTTTTGGAAAACTCAAAGGAATTCGCCAGTTAG 2860  
D 2888 CACATACTCCCGAGCATCTAGCTATGCTGGAGACCTTCGAGAACGCCCTCCCTGG 2947  
QY 2861 AAGCCAGAGACATGATAACGTGATCTTACAGATCTTTGACACAGTGAAGCATTA 2920  
D 2948 AAGCCACAGACATGATCAGGTGGTGTCTAAAGATGTTTAAATCAATTAAGAGATA 3007  
QY 2921 GAGAAAAATCTTTTAAATGTAGCAACCATCAGGATCGGAATCTTAAAGATGATG 2980  
D 3008 GAGAGTGTTCATCCAGCAGCCCTGTGGCC-----GAGGCGGAGGGAACCAATATGGAGG 3061  
QY 2981 AGATTACTCATGATAAAGAAAAAGCAGAACAAAAAGAAAGCTCAAGCTGTAGGCTAC 3040  
D 3062 AGAGCTCAAGAGACAAGACAAGCAGAGAGAGAAAGAAAGCCGAGATCGCCAGCTGC 3121  
QY 3041 ATGCCAGAGATCATGCTCAGATGCTGCTTACAGAAAAAATTCATTGAAACTCATA 3100  
D 3122 GCGGGAGAGATCATGSCCAGATGCTGAGATCGACGGGCACCTTCATTGACGAARACA 3181  
QY 3101 AACTCATGTATGACAATACATCAGAAATGCCCTGGGAAAGAAAGATTCCATTATGGAGGAG 3160  
D 3182 AAGAGCTCTTCCAGCAGACCCCTAGAGCTGGACACCTCTGCCCTCTGCCACTCT- - - - -TG 3235  
QY 3161 AGAGCAGCCAGCAGTCACTGACTCTAGAAATGCTTTGGGTCTTAAAGGGGTGCAT 3220  
D 3236 ACAGCAGCCCTCCGCTTTTCAGACGCAGCTCTTACAGCAGCTGGGCCAGCAGACACAGG 3295  
QY 3221 CTGTTACTGAAAGAGGTGCTGAGTGCATCCTTTTGCACAGAAAGAGAGGTGAA 3280  
D 3296 TCCCTGAACCCAGACAGTTGTCACTGTATATGTCAGAGAGGAGAGGTGACTG 3355  
QY 3281 TAGAAAAATATGCGCATATATTCGGCTGTGTCAGAAAAATCTACTGCTTAAACCCAGC 3340  
D 3356 TGGAGACGAGCGGATGCTTTGGCAGCGTTTGTTCAGAGGTCAACGGTCTGTCAAGAG 3415  
QY 3341 ACAGGGAAAAACCATAGAACTCTCAGGAGAGGCCCTAGACCCACTTTTCATGGATCCAG 3400  
D 3416 ACAGGACAAAAACCATCGGACCCAGAAAAA- - -TATGATCCATTATTCATGCACCCCG 3472  
QY 3401 ACTTGGCATATGGAACATTATACAGGAAGCTGTGTGATCATGTAATGCACGAGTGTCTGGC 3460  
D 3473 ATCTGTCTTGGGACACACACTGSCAGCTGTGGGCACGTTATGATGCCATTCCTTTTGGC 3532  
QY 3461 AGAAGTATTTGAAGCTCTAC- - - - -AGCTGAGCTCTCAGCAGCGCATTCATGTTG 3511  
D 3533 AAAGTATTTTGAATTCGTTCAAGCCAGGAGCAGCAGGAGGCGGCTGCGCTTGC 3592  
QY 3512 ACCTT- - -TTTGACTTGGAAAGTGAGAAATATCTTTGGCCCTCTTTGCAAAATCTCTGTGCA 3568  
D 3593 ACATAGCTACGATGATAGAAATGCGAGTTCTCTGCGCGCTCTGTGAGTGCCTGAGCA 3652  
QY 3569 ATACTGTGATCCCATTAATCTTTGCAACCTCAAAAGATAAACAGTGAAGATGAGATG 3628  
D 3653 ACAGGCTGATCCCC-----CTGCTGCTTCTCCCGAGGAGCATCTCAGCAGGAGGTTAA 3706  
QY 3629 CTCCTTGCTCACTTTTGACCTTGGCAGGTGGATACAGACTGTTCTTGGCCAGAAATCAG 3688  
D 3707 ATTTTTCAGACCAACCATGCTGGCACAGTGGACAGCAGTAAACAGCAGATATAAGG 3766  
QY 3689 GTTATAATATAAGACATCTCAAGGAGAAAAACCAATTCCTATTTTCTTAATCAAGGAA 3748  
D 3767 TGGTCCAGATGCTGAGGAGAAAGCAATGCTGCTGACACGCTTCTTTCAGAGGACACAG 3826  
QY 3749 TGGGAGATTTACTTTTGAGTTTCCATTCCTCCTGAGTTTGGCCTTGAGTCTTCGATTA 3808  
D 3827 AAGCCATGAATATAATACCGATCCCGAAGGCTTCAGGCCCTGATTTTATCCTAGGAACC 3886

QY 3809 AATATTCAATACATCAGGAATGGTTATTCTCTTTGCCACACAATTTATAGATTG 3868  
D 3887 CATATTCTGATAGCATAAAAGAAATGTTAAAGCATTTTGAAGGCTGCTTACAAGTGG 3946  
QY 3869 GATTGAAAGTCCACCTGATGAAAGGATCCTCGAGTCCCATCTGCTACCTCGAGCAGCT 3928  
D 3947 GACTGAAGGTTTATCTTAATGAAGTGCACCCCGTGTGCCATCTCTGCTGGGGACCT 4006  
QY 3929 GGGCTTTCACTATCCAGGCAATTTGAAATCTATTTGGGAGATGAAGGAAAACTCTGTTT 3988  
D 4007 GTGCATACACCATCCAGAGCATAGAAAGAAATTTTGAAGTATGAGGAGAAGCCTGTTT 4066  
QY 3989 GAGCACTTCAAAATAGCAGCATTAATGGTCTGAAGCATTAATGTCAGTTTGCATTTGCAC 4048  
D 4067 GACCTCTGCTTGTGACTGGAGACTGTTCAAGTCTGTTAAACAGGTTTTCAGCAGCAGC 4126  
QY 4049 AGAGGATTACCTGCTCCTGAGTCTGATACAGAAACATCTGGTTCGCTTCTATCAGTTG 4108  
D 4127 ATTGGACAGTGGGCTTACTTCTGTGTGATACAGACACTTCTGTAAACTCTTTGCATCCT 4186  
QY 4109 TTCTTCTTAAACATPAAATCAGAAATACACCATGCTTCTCTCTATAGATCTGTTTCA 4168  
D 4187 TGTGCTTAGTGACAGCTATGAAGACCTCCCGTGCATACATAGACATCGACATGTTCACT 4246  
QY 4169 TTTTGTGGGTGCTGTGTAGCATTTCCATCTTGTATTTGGGATGACCTGTTGATCTGC 4228  
D 4247 TGTGTGGGCTTGGTGTGCTTCCAGCTCTGCACTG- - - - -TCAGG 4291  
QY 4229 AGCCTTCTCAGTTAGTTCTTCTTATACACACCTTTATCTCTTCCATTTGATGATCACC 4288  
D 4292 ATTTTTCAGGAAGCAGCTGGCCACTGGGACCTGACATCTTCCACTTGGTTACCATGG 4351  
QY 4289 CACACATGCTTACAGTACTACTTACAGTACACACAGCCCTACC- - -CCTTCTCAGGTT 4345  
D 4352 CACACATGCTACAGATCTTACTTACCTCATGTCAGAGAGAAATGGCATGGATCAAGAGA 4411  
QY 4346 AAGAAGACAGTGAAGAGGCTCATTCGCGCATCTTCTTCTTTGCAAGAAATTTCTCAATATA 4405  
D 4412 ATCCCACTGGGAAGAAGAACTGGCCATCTCTCTTTGCAACAAAACCTTCAACAGTATA 4471  
QY 4406 CAAAGTGGCTCCATTTGGGTGTGATATTCCTGGTGGTATTTGTGGGTCTCACTGAGAA 4465  
D 4472 CTGGAAGTGCCTTTGAAAGAGAGCCCTCCGCTGGCAGCTGTGGAGGAGCGCTCGGCC 4531  
QY 4466 GCATCACCCCTTATCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4525  
D 4532 CCATCATCTTCTTCCCTCAAGTGTCTGCTTGTGTTTTTCCACTATTTAAATGGAGTCCGG 4591  
QY 4526 CGCTCAGGAACCTGCATACCAATTTCTGCAAGAGAGAGTACAGTGCATCTCTGAGTATC 4585  
D 4592 CCCCTCCAGACCTTCA- - -AGTTTCTGGAACAAGCCATTTTGAACACTTATGTAACATCC 4648  
QY 4586 TATCTTTTACTACAAATTTGTTCTGCTCTTCCAGGAATATTTGGGATCTGTAAGGCGCT 4645  
D 4649 TTTCCCTACCAACCAACCTCATTCACCTTTTCAAGAAAAACAGTGACATCATGAACCTCC 4708  
QY 4646 TGCTCAGAGCGGTGTGCAGATCTGCTTACTAACTGTTTGAAGCAAAAAACACCG 4705  
D 4709 TGATTGAAATTTGGTGCCAGAACAGTGAAGTTAAACGGTATCTAAATTTGGCGAGAGAGAG 4768  
QY 4706 TGGTCAAGTACCCCTAGAAAAAGAAATAGTTTGTATAGAGCTTCTCTGATGACTATAGCTGCC 4765  
D 4769 CGATAGCTTACCCAGAGGAGCTAACAACTGATAGACCTTCCAGAGGATTTACAGAGCC 4828  
QY 4766 TCTGAATCAAGTCTCTCATTTTCCAGGTGCCAGCGTCTGCAGATGTAGCGAAAGCATC 4825  
D 4829 TCATTAAACCAAGCATCCAATTTCTCGTGGCCCCAAATCAGGTGGCGACAGAGCAGCTC 4888  
QY 4826 CTGCTCTGCGCTTTTCTGTGGGGCTATATGTTCTCAGAACATTTGCTGCCAGGAAA 4885  
D 4889 CTACTCTGCTGCTGTGTGGAGTCTCTCTGCTCTCAGATTTACTGCTGCCAAGCTG 4948  
QY 4886 TTGTGAACGGGGAAGAGTTGGAGCTTGCATTTTTCAGGCATTTCTCAGCAGCTTCACTGTGGACCGGAG 4945

```
Db 4949 ACCTGGAGGTGAGGACGTCGGAGCTGCACAGCACACACTACTCTCGGCTCGGGG 5008
QY 4946 TCTGCATTTTCCATAAAATCAGAGATGCCGAGTGGTCTGTTGAAGGTAAAGCCAGAG 5005
Db 5009 CCGGCATCTTCTGAGAGTGGGGAATGTCAAGTGTATTTTGTAGTGGCAAAACCAAG 5068
QY 5006 GCTGTGCCATATCCAGCTCTTACTTGGATGAATATGGAGAAACAGACCCCTGGCTGAAG 5065
Db 5069 GATGTTTATTTTCTCTCTTACCTTGACGACTATGGAGAGACCGACCGAGCTCAGAC 5128
QY 5066 GGGCAACCCCTTCAATTTATCTCGTACGCGTATCGGAAGCTTCCATTGTTGCTGCGCAAC 5125
Db 5129 GAGGAATCTTTACATTTATGCGAAGAGCGGTTTCCGAAGATCCGAAGCTCTGCGACG 5188
QY 5126 AACACTGCTATTAGAAAGAGATTTGCTTAGGAGCCAGAGACTAATCAGATGTTTATTTGGAT 5185
Db 5189 AGCATAGTATCAGAGGAGATCGGACACGCGCAGGAGGCTAACCCAGAGCCCTGGTCGAA 5248
QY 5186 TCAACTGGCAG 5196
Db 5249 TTGACTGGCAG 5259

RESULT 12
AR030785
LOCUS AR030785 1001 bp DNA linear PAT 29-SEP-1999
DEFINITION Sequence 2 from patent US 5861312.
ACCESSION AR030785
VERSION AR030785.1 GI:5943999
KEYWORDS
SOURCE
ORGANISM
REFERENCE
  1 (bases 1 to 1001)
AUTHORS Varshavsky,A. and Kwon,Y.Tae.
TITLE Nucleic acid encoding mammalian Ubr1
JOURNAL Patent: US 5861312-A 2 19-JAN-1999;
FEATURES
  Location/Qualifiers
    source
      1..1001
        /organism="unknown"
BASE COUNT 363 a 186 c 205 g 247 t
ORIGIN
```

```
Query Match 19.1%; Score 996.2; DB 6; Length 1001;
Best Local Similarity 99.7%; Pred. No. 4.9e-232;
Matches 998; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2059 ATGGATCCCAATAAGTCTTGTGTACTTGGTACTTTCAGAGGTATGAACCTGCGGAGGCTTTT 2118
Db 1 ATGGATCCCAATAAGTCTTGTGTACTTGGTACTTTCAGAGGTATGAACCTGCGGAGGCTTTT 60
QY 2119 AACAAACCATATCTACAAAAGACCAGGATTTGATTAACAATATAACACTAATAGAA 2178
Db 61 AACAAACCATATCTACAAAAGACCAGGATTTGATTAACAATATAACACTAATAGAA 120
QY 2179 GAATGCTTCAGTCTCTCATCTATTTGTGGTGAGCTTATGCTTACTTGGAGTGGGAAT 2238
Db 121 GAATGCTTCAGTCTCTCATCTATATTTGGTGAGCTTATGCTTACTTGGAGTGGGAAT 180
QY 2239 GTGACCAAAAGAGAGTCAATGAGAGAAATCATTTCACTTGTCTTTGCAATGAACCCATG 2298
Db 181 GTGACCAAAAGAGAGTCAATGAGAGAAATCATTTCACTTGTCTTTGCAATGAACCCATG 240
QY 2299 CCACACAGTCCCATTTGCCAAAATTTTACCTGAGAAATGAATAATGAACCTGGCTTAGAG 2358
Db 241 CCACACAGTCCCATTTGCCAAAATTTTACCTGAGAAATGAATAATGAACCTGGCTTAGAG 300
QY 2359 AATGTCTATAAACAAGTGGCCACATTTAAGAAACCAAGGTGTATCAGGCCATGAGTTTAT 2418
Db 301 AATGTCTATAAACAAGTGGCCACATTTAAGAAACCAAGGTGTATCAGGCCATGAGTTTAT 360
QY 2419 GAACTAAAAGATGAATCACTGAAAGACTTCAATATGTACTTTTATCATTTACTCCAAAACC 2478
```

```
Db 361 GACTAAAAGATGAATCACTGAAAGACTTCAATATGTACTTTTATCATTTACTCCAAAACC 420
QY 2479 CAGCATAGCAAGGCTGAACATATGCAAGAAAAAGAGAAAAACAAGAAAAACAAGATGAA 2538
Db 421 CAGCATAGCAAGGCTGAACATATGCAAGAAAAAGAGAAAAACAAGAAAAACAAGATGAA 480
QY 2539 GCATTTGCCGCCACACACACCTCTCTGAATTTCTGCCCTGCTTTTCAGCAAAAGTGAATTAACCTT 2598
Db 481 GCATTTGCCGCCACACACACCTCTCTGAATTTCTGCCCTGCTTTTCAGCAAAAGTGAATTAACCTT 540
QY 2599 CTCAACTGTGATATCATGATGATCATCTCTCAGGACCGTATTTGAGCGGGCAATTAACACA 2658
Db 541 CTCAACTGTGATATCATGATGATCATCTCTCAGGACCGTATTTGAGCGGGCAATTAACACA 600
QY 2659 GATTCTAACTTTGGGACCCGAAGGATGCTCCAAATGGCTTTTCATATTTCTGGCATTTGGGT 2718
Db 601 GATTCTAACTTTGGGACCCGAAGGATGCTCCAAATGGCTTTTCATATTTCTGGCATTTGGGT 660
QY 2719 TTACTAGAAGAGAACACAGCTTCAAAAAGCTCCTCGAAGAGTAAAGTAAACATTTTGACATTT 2778
Db 661 TTACTAGAAGAGAACACAGCTTCAAAAAGCTCCTCGAAGAGTAAAGTAAACATTTTGACATTT 720
QY 2779 TATCATTAAGGCTTCAAGATTTGGGAAGTTTCAAGGATGCTCCAAATGGCTTTTCATATTTCTGGCATTT 2838
Db 721 TATCATTAAGGCTTCAAGATTTGGGAAGTTTCAAGGATGCTCCAAATGGCTTTTCATATTTCTGGCATTT 780
QY 2839 CTCAAAAGGAATTTCCCAAGTTAGAAAGGCCAAGGACATGATAAGCTGGATCTTCAGATG 2898
Db 781 CTCAAAAGGAATTTCCCAAGTTAGAAAGGCCAAGGACATGATAAGCTGGATCTTCAGATG 840
QY 2899 TTTGACACAGTGAAGCGGATTAAGAGAAAAATCTTTGTTTAAATTGTTAGCAACCATCAGGA 2958
Db 841 TTTGACACAGTGAAGCGGATTAAGAGAAAAATCTTTGTTTAAATTGTTAGCAACCATCAGGA 900
QY 2959 TCGGAATCTATTAGAATGATGAGATGATGAGATGATGAGATGATGAGATGATGAGATGATGAGATG 3018
Db 901 TCGGAATCTATTAGAATGATGAGATGATGAGATGATGAGATGATGAGATGATGAGATGATGAGATG 960
QY 3019 AAGCTTGAAGCTGTAGCTACATCGCCAGAAAGATCATGGC 3059
Db 961 AAGCTTGAAGCTGTAGCTACATCGCCAGAAAGATCATGGC 1001

RESULT 13
AR121464
LOCUS AR121464 1001 bp DNA linear PAT 16-MAY-2001
DEFINITION Sequence 2 from patent US 6159732.
ACCESSION AR121464
VERSION AR121464.1 GI:14105040
KEYWORDS
SOURCE
ORGANISM
REFERENCE
  1 (bases 1 to 1001)
AUTHORS Varshavsky,A. and Kwon,Y.Tae.
TITLE Nucleic acid encoding mammalian Ubr1
JOURNAL Patent: US 6159732-A 2 12-DEC-2000;
FEATURES
  Location/Qualifiers
    source
      1..1001
        /organism="unknown"
BASE COUNT 363 a 186 c 205 g 247 t
ORIGIN

Query Match 19.1%; Score 996.2; DB 6; Length 1001;
Best Local Similarity 99.7%; Pred. No. 4.9e-232;
Matches 998; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2059 ATGGATCCCAATAAGTCTTGTGTACTTGGTACTTTCAGAGGTATGAACCTTGGCAGGCTTTT 2118
Db 1 ATGGATCCCAATAAGTCTTGTGTACTTGGTACTTTCAGAGGTATGAACCTTGGCAGGCTTTT 60
QY 2119 AACAAACCATATCTACAAAAGACCAGGATTTGATTAACAATATAACACTAATAGAA 2178
```





Db	361	GAACATAAAGATGAATCACTGAAGACITTCATAATATGTACTTTTATCATTTACTTCCAAAACC	420
Qy	2479	CAGCATAGCAAGGCTGAACATATGCAGAGAAAAGAGAAAACAAGATGAA	2538
Db	421	CAGCATAGCAAGGCTGAACATATGCAGAGAAAAGAGAAAACAAGATGAA	480
Qy	2539	GCATTGCCGCCACCAACCACTCCTCTGAATTCCTCCCTGCTTTTCAGCAAAAGTGATTAACCTT	2598
Db	481	GCATTGCCGCCACCAACCACTCCTCTGAATTCCTCCCTGCTTTTCAGCAAAAGTGATTAACCTT	540
Qy	2599	CTCAACTGTGTATCATGTATGATCTCTCAGGACCGTATTTGAGCGGCAATAGACACA	2658
Db	541	CTCAACTGTGTATCATGTATGATCTCTCAGGACCGTATTTGAGCGGCAATAGACACA	600
Qy	2659	GATTTCTAAGTGTGGACCGAAGGATGCTCCAAATGGCTTTTCATATTCGCGATTTGGGT	2718
Db	601	GATTTCTAAGTGTGGACCGAAGGATGCTCCAAATGGCTTTTCATATTCGCGATTTGGGT	660
Qy	2719	TTACTAGAGAGAAGCAACAGCTTCAAAAAGCTCTCGAAGAAGAAGTAACATTTGACTTT	2778
Db	661	TTACTAGAGAGAAGCAACAGCTTCAAAAAGCTCTCGAAGAAGAAGTAACATTTGACTTT	720
Qy	2779	TATCATAAGGCTTCAAGATTTGGAAGTTTCAGCCATGAATATACAAATGCTTTTGGAAAAA	2838
Db	721	TATCATAAGGCTTCAAGATTTGGAAGTTTCAGCCATGAATATACAAATGCTTTTGGAAAAA	780
Qy	2839	CTCAAGGAATTTCCCGAGTTAGAAGCCAGAGGACATGATTAACGTTGGTACTTTCAGATG	2898
Db	781	CTCAAGGAATTTCCCGAGTTAGAAGCCAGAGGACATGATTAACGTTGGTACTTTCAGATG	840
Qy	2899	TTTGACACAGCTGAAGCGATTAGAGAAAAATCTTGTTTTAAATGTAGCAACCATCAGGA	2958
Db	841	TTTGACACAGCTGAAGCGATTAGAGAAAAATCTTGTTTTAAATGTAGCAACCATCAGGA	900
Qy	2959	TCGGAATCTATTAAAGATGATGAGATTACTTCATGATAAAGAAAAAGCAGAACGAAAAAGA	3018
Db	901	TCGGAATCTATTAAAGATGATGAGATTACTTCATGATAAAGAAAAAGCAGAACGAAAAAGA	960
Qy	3019	AAAGCTGAAGCTGCTAGGCTACATCGCCAGAGATCATG	3057
Db	961	AAAGCTGAAGCTGCTAGGCTTCATCGCCAGAGATCATG	999
RESULT 15	AB002347	6158 bp	mrna linear PRI 06-OCT-2001
LOCUS	Human mRNA for KIAA0349 gene, partial cds.		
DEFINITION	AB002347		
ACCESSION	AB002347.1 GI:2224638		
VERSION	KIAA0349.		
KEYWORDS	Homo sapiens (human)		
SOURCE	Homo sapiens		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
REFERENCE	1		
AUTHORS	Nagase, T., Ishikawa, K., Nakajima, D., Ohira, M., Seki, N., Miyajima, N., Tanaka, A., Kotani, H., Nomura, N. and Ohara, O.		
TITLE	Prediction of the coding sequences of unidentified human genes. VII. The complete sequences of 100 new cDNA clones from brain which can code for large proteins in vitro		
JOURNAL	DNA Res. 4 (2), 141-150 (1997)		
MEDLINE	97349984		
PUBMED	9205841		
REFERENCE	2 (bases 1 to 6158)		
AUTHORS	Ohara, O., Nagase, T., Kikuno, R. and Nomura, N.		
TITLE	Direct Submission		
JOURNAL	Submitted (28-MAR-1997) Osamu Ohara, Kazusa DNA Research Institute; 1532-3, Yana, Kisarazu, Chiba 292-0812, Japan		
FEATURES	(E-mail:cdnainfo@kazusa.or.jp, Tel:+81-438-52-3913)		
source	Location/Qualifiers		
	1..6158		
	/organism="Homo sapiens"		
	/mol_type="mrna"		

gene	/db_xref="taxon:9606"				
	/clone="HG1561"				
	/sex="male"				
	/tissue_type="brain"				
	/clone_lib="pBluescriptII SK plus"				
	1..6158				
	/gene="KIAA0349"				
CDS	<1..3828				
	/gene="KIAA0349"				
	/codon_start=1				
	/protein_id="BAA20806.1"				
	/db_xref="GI:2224639"				
	/translation="LKVLISKPTPEWSDLELRQKLEGFDAFLLELLKCMQMDPIITROV GOHMEPEWEAAFTLQMLKLVTHVISMQDMCASDEKVLIEAYKKCLAVLMQCHGYTD GEQITLISICGSHVETIRYCVQKQVSHLPVSRLLAGLVLLSKSEVAYVAFPELLPL SELSPLMLIEHPLRLCLVCAQVHAGWMRRNGFSLVNQIYYIYHNKVRREMFQDVVML QTGVSMMDPNHFLIMLSRPELYQIFSTPDYQKFRFSSEITHKDVQVQNNLTIEMLYL IIMLGERFSPGVQVNVATDEIKREIIHQLSIKPMHSELVKSLPEDEKNTKGESVI EVAHFKKPGLTGRGMYELKPECAKEENLYFYHFSRAEQSKAREAQRLKLRQNRDPA LPPVLPFCPLFASLVNLIQSDVLMCIIMGTILQWAVEHNGYAMSESMQRLVHLIGM ALQEKQHLNVTEEHVVFTFTQKISKPEAGPKNSPILAMLETKLQNAIARLRKIM IRWLKTFNAVKKHMRSSPTPVAETEGTMEESSRDKAERKKAIEARLRKIM AOMSEMRHFIDENKELFQOTLELDASTAVLDHSPVASDMTLTALQPTQVPEQORQ FVTCILCOEEQEVKVESRAVLAAFVORSTVLSKNRSKFIDQPEKYDPLFPHPLDLSCG THTSSCGHIMHACWQRYFDSVQAKQRRQORLRLHTSYDVENGEFLCPLCECLNTV IPLLPRNIFNNKLNESDQPNLTQWRTISQIKALQFLRKEESTPNNASTKNSNV DELQPEGRFPDRPKIPYSESIKEMLTTFGTATYKVLKVHPNEEDRPVIMCMGSC AYTQSTIERISLDEKPLFGPLPCRLDDCLSLRFAAAHMTVASVSVVQHFCKLFA SLVNDPSHEELPCITLIDMFHLVGLVLAFLALQCOQPSGISLGTDLHILFHVMTMAH IIQILLTSCTEENGMDQENPPCEESAVLALYKTLHOYTGSALKEIIPSGWHLRWSRA GIMPFKCSALFPHYLVGVSPDPPIQVPTGTFHFEHLCSYLSLPNNLILCLFOENSEIMN SLTESWCNREVRKYLEGKDAIRYPRESNKLINLPEDYSLINQASNFSPCKSGGDK SRAPTLCLVCGSLCSYCCYCCQTEGEDGACATHTYSCGSGVGIFLVRRECVQLFL AGTKGCFYSPPLYDDYGETDQGLRRGNPLHLCKERFKIKQLKWHQHSVTEIQAHOE ANOTLVGIDWQHL"				
BASE COUNT	1832 a	1199 c	1320 g	1807 t	
ORIGIN					
	Query Match	18.8%;	Score 979.4;	DB 9;	Length 6158;
	Best Local Similarity	55.9%;	Pred. No. 6.1e-228;		
	Matches 2156;	Conservative	0;	Mismatches 1611;	Indels 90; Gaps 12;
QY	1390	CTAAGTATATCTCGATCAGCAAAACCACATATATGGACAGAAAGATTAAAGATGCAGTTC	1449		
Db	1	CTCAAGTATGTGTTAATTAGCAAACTCACTGAATGGTCAGATGAGCTGAGCGAGAAGTTC	60		
QY	1450	CTTGAAGGTTTTCGATCTTTTTGAAGATCTTTTACCTGTATGTCAGGGAATGGAAGAAATC	1509		
Db	61	CTAGAAGGTTTGTATGCTCTTTTGGAAATTAATAATGTATGCAGGGAATGGATCCAATTT	120		
QY	1510	CGAAGACAGTTGGGCAACACATCTGAAGTGGATCCTGATTTGGGAGGCTGCCATTGCTATA	1569		
Db	121	ACAGCTCAAGTAGGACACACATATTGAATGGAACACAGAGTGGGAAGCAGGCTTCACACTA	180		
QY	1570	CAGATGCAATTGAAGAATATTTTACTCATGTTCCTCAAGAGTGGTGTGCTGTGTGATGAAGAA	1629		
Db	181	CAAAATGAATTAACACATGTCATTTCAATGATGCAGGACTGGTGTGCTTCAGATGAAGAAA	240		
QY	1630	CTCTTACTTGGGCTTATAAGAATGTCAACAAGCTGTGATGAGGTGCAGTACCAAGTTTC	1689		
Db	241	GTGTTAATCGAAGCTTACAGAAATGTCTCGCTGCTACTGATGCAGTGTCTATGGTGGTTAT	300		
QY	1690	ATATCTAGTAGCAAGACAGT---AGTACAATCGTGTGGACATAGTTTGGAAAAAAGATCC	1746		
Db	301	ACTGATGGTGAACAGCCAAATCACACTAAGCATTTGTGGACATTCAGTGGAAACTATCAGA	360		
QY	1747	TACAGACTATCTGAGGATCTTGTGAAGCATACATCTGCCACTCTCTAGGACCCCTTGCTGGT	1806		
Db	361	TACTGTGTTTCCCAAGAAAAAGTTAGCATTCACCTCCCAAGTTTCTCGCTTACTTTCAGAGT	420		
QY	1807	CTTCATGTACGTTTTAAGCAGCGCTGGGTGCTGCTTTTCAAGACTGCATGAATTTGTCTCTTT	1866		
Db	421	TTACATGTATTATTAAAGCAAAAGTGAAGTGGCATATAAATTTCCAGAGCTCCTACCTCTA	480		

OY 1867 GAGACTTCAAGTAGAGTACTAGTGGAAATCCTTTACGTTGCTGTTGGTCC 1926  
 DB 481 AGTGAACCTTAGCCGCCACCCATGTTGATGAAACCCCTTAGATGTCGTTCTGTGTC 540  
 OY 1927 CAGTGTGTTGCTGAGATGTGGCGAAGAAATGAGACTGCTCTTATTTAGCCAGGCTTTAT 1986  
 DB 541 CAAGTACATGCCGGAATGTGAGAAAGAAATGGGTTCTCTAGTAACACAGATTTATTAC 600  
 OY 1987 TACCAAGATGTTAATGTCAGAGAAAGAAATGTAATGAATAATTCATTCCTTCAGATT 2046  
 DB 601 TACCAATATGTAATAATGTCAGAGTATGTTGACAAAGATGTAATGTAATCTTCAGACA 660  
 OY 2047 GGTGATCTTTAATGATGCCATTAAGTCTGTTACTGTTACTGTTACTGAGATGAGT 2106  
 DB 661 GGTGCTCCATGATGATGATCCAAATCATTTCTGATGATCATGCTGACGCCGTTTGACTT 720  
 OY 2107 GCCGAGGCTTTTACAGACATATCTACAAA-----GACGAC 2145  
 DB 721 TATCAGATTTTCAAGTACCCAGATATGAAAAAGATTAGTTCTGAGATTACCCATTAA 780  
 OY 2146 GATTGATTTAAACAATATATATCACTAATAGAGAAATGCTTCAGGTTCTCATTTAAT 2205  
 DB 781 GATGTTGTTACAGCAAGACAACTACTCTAATAGAAATGCTTACTCATTTAATAGCTT 840  
 OY 2206 GTGGGTGAGCGTTATGTAACCTGAGTGGAAATGTGACCAAGAGAGTCAATGACA 2265  
 DB 841 GTTGAGAGAGATTTAGTCTGAGAGTGGACAGTAATGCTACAGATGAAATCAAGCCA 900  
 OY 2266 GAAATCATTCACCTGCTTTGCTGATTAACCCATGCCACAGAGTCCATGCCAAAATTTA 2325  
 DB 901 GAGATTAATCCATCACTTATGATGATCAAGCCTATGCTCATATGTAATGTTAAAGCTTTA 960  
 OY 2326 CCGTGAATGAAATATATGAACTGCTTGAAGATGTCATTAACAAAGTGGCCACATT 2385  
 DB 961 CCGTGAATGAAATATATGAACTGCTTGAAGATGTCATTAACAAAGTGGCCACATT 2385  
 OY 2386 AAGAAACAGAGTATCAAGGATGAGGATTTATGAATTTAAAGATGATCACTGAAAGAC 2445  
 DB 1021 AAGAAACAGAGTATCAAGGATGAGGATTTATGAATTTAAAGATGATCACTGAAAGAC 2445  
 OY 2446 TTCAATATGACTTTTATCATTTACTTACCAAAACCCAGCATGAGCAAGGCTTACATATGAC 2505  
 DB 1081 TTCAATATGACTTTTATCATTTACTTACCAAAACCCAGCATGAGCAAGGCTTACATATGAC 2505  
 OY 2506 AAGAAACAGAGTATCAAGGATGAGGATTTATGAATTTAAAGATGATCACTGAAAGAC 2565  
 DB 1141 CGAAATTTGAAAGCAAAATATGAGATATACAGCACTCCACGCTGGTGTTCCTCCA 1200  
 OY 2566 TTCTGCCCTGCTTTCAGCAAAAGTATTAACCTTCAACTGTGATATCATGATGAT 2625  
 DB 1201 TTCTGCCCTGCTTTCAGCAAAAGTATTAACCTTCAACTGTGATATCATGATGAT 2625  
 OY 2626 CTCAGAGCCGATTTGAGCGGCATATACACAGATTTCACTTGTGACCGCAAGGATG 2685  
 DB 1261 ATGGGAACAAATTCGCAATGGGCTGTGGAACATTAATGATGATGCTGTGCAAGTCCAG 1320  
 OY 2686 CTCCAATATGCTTTTATCATTTCTGCAATGGGTTACTAGAGAGAGCAAGCAAGCTTCA 2745  
 DB 1321 CTGCAAAAGGTTGATTAATTTGATGATGATGATGATGATGATGATGATGATGAT 1380  
 OY 2746 AAGCTCTGAG---AAGAAATGATGATGATGATGATGATGATGATGATGATGAT 2802  
 DB 1381 AATGTCAGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1440  
 OY 2803 AGTTCAGCATGAAT-----ATCAAAATGCTTTGAGAAAACTCAAGAAATTC 2853  
 DB 1441 GAAGCGCCAAAATTTCTCTAGCATGATGATGATGATGATGATGATGATGATGAT 1500  
 OY 2854 CAGTTAGAGGCGAAGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 2913  
 DB 1501 TACCTAGAGTCCAAAGAGATGATGATGATGATGATGATGATGATGATGATGAT 1560

OY 2914 CGATTAGAGAAAATCTGTTTAATGTTAGAACACCATCAGATGCGAATCTATTAG 2973  
 DB 1561 AAGATGAGAGAGA-----GTTCACTACAGAGTCCCGGAGAGACAGAGAAACATA 1614  
 OY 2974 AATGATGAGATTTACTGATTAAGAAAGAAACAGACCAAGAAAGAAAGAGTCACTCT 3033  
 DB 1615 ATGGAAGAGATTTCAAGGACAAAGACAAACCTGAGAGAAAGAAAGAGATTTGCC 1674  
 OY 3034 AGGCTACATGCGCCAGAGATCAGTGGTACATGTTGCTGCTTACAGAAAACTCATGAA 3093  
 DB 1675 AGACTGGCAGAGAAAGATCATGCTCAGATGCTGAAATGACAGCGCATTTATGAT 1734  
 OY 3094 ACTCATTAATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3153  
 DB 1735 GAAACCAAAAGACTTTTACAGACATTAATGATGATGATGATGATGATGATGATGAT 1794  
 OY 3154 GAGAGAGAGACACCCAGCAGAGTGTGAGTACTGATGATGATGATGATGATGATGAT 3213  
 DB 1795 GATCA-----TAGCCCTGTGGCTTACATGATGATGATGATGATGATGATGATGAT 1848  
 OY 3214 GGTCCATCTGTTACTGAAAGAGAGTGTGATGATGATGATGATGATGATGATGATGAT 3273  
 DB 1849 ACTCAGGTTCTGAAACAAAGACATTTGATGATGATGATGATGATGATGATGATGAT 1908  
 OY 3274 GTGAAATTAAGAAATATGATGATGATGATGATGATGATGATGATGATGATGAT 3333  
 DB 1909 GTTAAAGTGAAGAGCAGGAGGATGCTTGGCAGCATTTGTTCAAGATGATGATGAT 1968  
 OY 3334 ACCAGACAGAGGGAAGAACCATGATGATGATGATGATGATGATGATGATGATGATGAT 3393  
 DB 1969 TCAAAAACAGAAATTAATTTATTCAGATGATGATGATGATGATGATGATGATGAT 2025  
 OY 3394 GATCAGACTTGGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3453  
 DB 2026 CACCTGATGCTGCTTGTGGAACACACATGATGATGATGATGATGATGATGATGATGAT 2085  
 OY 3454 TGTGCGAGAAATTTTGAAGTGTGATGATGATGATGATGATGATGATGATGATGATGAT 3513  
 DB 2086 TGTGCGAGAAATTTTGAAGTGTGATGATGATGATGATGATGATGATGATGATGATGAT 2145  
 OY 3514 CTTT-----TGACTTGGAAAGTGAAGATATCTTGGCCCTTGGCAATCT 3561  
 DB 2146 CGCTTACATGAGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2205  
 OY 3562 CTGTGCAATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3621  
 DB 2206 TTGAGTAATCTGTTATTC-----TCTGTCTCTCCAGAAATTTTAAACAC 2259  
 OY 3622 GCAGATGCTTGTCTCAACTTTTACCCCTGGCAGGTGATGATGATGATGATGATGATGAT 3681  
 DB 2260 AGGTTAAATTTTTCAGACCAACCAATCTGATGATGATGATGATGATGATGATGATGAT 2319  
 OY 3682 ATATCAGGTTATATATTAACATGATGATGATGATGATGATGATGATGATGATGATGAT 3741  
 DB 2320 ATTAAGACATTAAGTTTCTTAGGAAAGAAAGTCTCTTAATTAAGTCTTACAAAG 2379  
 OY 3742 CAAGGATGAGATTTACTTGTGAGTGTGATGATGATGATGATGATGATGATGATGATGAT 3801  
 DB 2380 AATTCAGAAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2439  
 OY 3802 TCGATTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3861  
 DB 2440 AAGATCCCTTATCTGAGACATTAAGAAATGATGATGATGATGATGATGATGATGATGAT 2499  
 OY 3862 AGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3921  
 DB 2500 AAGTGGGACTTAAGGTTTCAATGATGATGATGATGATGATGATGATGATGATGATGAT 2559  
 OY 3922 AGCAGCTGGCTTCTCTATCCAGGCAATGAAATATTAATGAGGAAAGAAACCT 3981  
 DB 2560 GGTAGCTGGCGTACACATTCACCAAGCATGAAAGAAATTTGATGATGATGATGATGAT 2619  
 OY 3982 CTGTTTGAGACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4041



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 26, 2003, 19:32:57 ; Search time 820.105 Seconds

(without alignments)  
17132.652 Million cell updates/sec

Title: US-09-724-126A-18

Sequence: 1 atggcgagagagagagctgtg.....tcaactgagcagctactgtga 5205

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq\_19Jun03:\*  
1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1880.DAT:\*  
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1881.DAT:\*  
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1882.DAT:\*  
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1883.DAT:\*  
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1884.DAT:\*  
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1885.DAT:\*  
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1886.DAT:\*  
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1887.DAT:\*  
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1888.DAT:\*  
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:\*  
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:\*  
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:\*  
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:\*  
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:\*  
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:\*  
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:\*  
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:\*  
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:\*  
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:\*  
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:\*  
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:\*  
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:\*  
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:\*  
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:\*  
25: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	5115	98.3	7742	25	ABZ24689	Human cell growth, CDNA encoding mouse
2	4085.6	78.5	6395	20	AAV99308	Nucleotide sequence
3	4085.6	78.5	6395	22	AAV99308	Human CDNA sequence
4	2434.4	46.8	2550	22	ABQ14878	Human ubiqutin re
5	1311	25.2	1635	24	ABQ75898	Human polynucleoti
6	1097.8	21.1	6840	22	AAK51709	Human polynucleoti
7	1088.4	20.9	6850	22	AAK52693	Human polynucleoti
8	996.2	19.1	1001	20	AAK03300	Partial CDNA encod

9	996.2	19.1	1001	22	AAK66934	Nucleotide sequenc
10	812	15.6	3327	20	AAK35730	CDNA encoding a pr
11	812	15.6	3302	20	AAK35731	CDNA encoding a pr
12	652.8	12.5	733	21	AAA02411	Human colon cancer
13	634.6	12.2	756	21	AAA02327	Human colon cancer
14	534.6	10.3	712	22	AAH07621	Human CDNA clone (
15	450.8	8.7	777	22	AAH07068	DNA encoding novel
16	429	8.2	1464	23	AAH70104	DNA encoding novel
17	373.2	7.2	1772	24	AAH90577	Human polynucleoti
18	271.4	5.2	1172	23	AAH70851	DNA encoding novel
19	267	5.1	2616	24	AAH70161	Fibrinogen 9.57 co
20	249	4.8	972	24	AAK35394	Human CDNA encodin
21	246.8	4.7	807	20	AAK36891	Gastric cancer ass
22	238.4	4.6	264	21	AAK0233	Human colon cancer
23	218.2	4.2	4573	25	ACA03886	CDNA upregulated i
24	195	3.7	6278	23	ABL14799	Drosophila melanog
25	191.6	3.7	800	20	AAK39892	Gastric cancer ass
26	159	3.1	455	20	AAH89674	EST clone DA490.
27	142	2.7	643	24	ABL89546	Human polynucleoti
28	132	2.5	13134	23	ABL14798	Drosophila melanog
29	128	2.5	529	22	AAK62305	Human immune/haema
30	116.6	2.2	308	16	AAH21876	Human gene signatu
31	104.8	2.0	507	23	ABV59104	Human prostate exp
32	100.6	1.9	1382	25	ABO82975	Human lung specific
33	71	1.4	7428	22	ABA07164	Human pancreatic c
34	71	1.4	7428	25	ABT16907	Human secreted pro
35	71	1.4	7428	25	ABZ67490	Human secreted pro
36	66	1.3	8676	22	ABA07159	Human pancreatic c
37	66	1.3	31931	22	ABA07160	Human pancreatic c
38	66	1.3	51961	25	ABT16905	Human secreted pro
39	66	1.3	51961	25	ABZ67488	Human secreted pro
40	60	1.2	60	24	ABN36159	Human spliced tran
41	54.8	1.1	2793	21	AAA70146	Plasmodium falcipa
42	49.4	0.9	12409	24	AAH63313	Chemically pretrea
43	49.2	0.9	1727	22	AAH17089	Human CDNA sequenc
44	48.6	0.9	4590	22	AAH24065	Yeast AOD9604-asso
45	46.2	0.9	581	22	AAH09232	Human CDNA clone (

ALIGNMENTS

RESULT 1  
ID ABZ24689 standard; CDNA: 7742 BP.

XX ABZ24689;

DT 07-APR-2003 (first entry)

DE Human cell growth, differentiation and death protein CGDD-1 cDNA.

XX CGDD-1, cell growth; cell differentiation; cell death; human;  
XX cytosolic; antiarteriosclerotic; hepatotropic; antiinflammatory;  
XX antiparasitic; antianaemic; ophthalmological; auditory;  
XX anticonvulsant; cerebroprotective; nootropic; neuroprotective;  
XX antiparkinsonian; neuroleptic; tranquilizer; immunosuppressive;  
XX anti-HIV; antiallergic; antiasthmatic; antitumor; antidiabetic;  
XX dermatological; nephrotoxic; antirheumatic; antifungal;  
XX antitumor; vulnery; virucide; antibacterial; fungicide;  
XX antiparasitic; protozoacide; antihelminthic; antifertility;  
XX gynaecological; ubiquitin protein ligase; enzyme; gene therapy;  
KW microarray; gene; ss.

OS Homo sapiens.

XX Key Location/Qualifiers

FT CDS 13..5229

FT /\*tag= a

PN /product= "Human CGDD-1"

WO200297032-A2.

PD 05-DEC-2002.  
XX 05-APR-2002; 2002WO-US11152.  
XX 06-APR-2001; 2001US-282110P.  
PR 11-APR-2001; 2001US-283294P.  
PR 26-APR-2001; 2001US-286820P.  
PR 27-APR-2001; 2001US-287228P.  
PR 18-MAY-2001; 2001US-291662P.  
PR 18-MAY-2001; 2001US-291846P.  
PR 25-MAY-2001; 2001US-293727P.  
PR 01-JUN-2001; 2001US-295263P.  
PR 01-JUN-2001; 2001US-295340P.  
PR 15-JAN-2002; 2002US-349705P.  
XX (INCY-) INCYTE GENOMICS INC.  
XX Azimzai Y, Au-Young JK, Batra S, Baughn MR, Becha SD, Borowsky ML;  
PI Burford N, Ding L, Elliott VS, Emerling BM, Gandhl AR, Gietzen KJ;  
PI Griffin JA, Hafalia AJA, Honchell CD, Lal PG, Lee SY, Lu DAM;  
PI Arvizu CS, Ramkumar J, Reddy R, Sanjanwala MM, Tang YT, Walla NK;  
PI Wang YE, Warren BA, Xu Y, Yang J, Yao MG, Yue H, Zebajadian Y;  
XX WPI: 2003-140453/13.  
DR P-PSDB; ABP58330.  
XX Novel human proteins associated with cell growth, differentiation and  
PT death, useful for treating, diagnosing or preventing cancer,  
PT developmental, neurological, reproductive or autoimmune/inflammatory  
PT disorders -  
XX Claim 5; Page 216-218; 238pp; English.  
XX The present sequence is that of Incyte clone 1351608CBI encoding  
CC human CGDD-1, a novel protein associated with cell growth,  
CC differentiation and death. A representative cDNA library for the  
CC polynucleotide is PGAN0701 from paraganliomic tumour tissue.  
CC Structural features establish the encoded protein as being  
CC associated with cell growth, differentiation and death, with  
CC further evidence suggesting it to be a ubiquitin protein ligase.  
CC The invention is based on novel human CGDD-1 to -21 proteins (see  
CC ABP58330-50), the polynucleotides encoding them (see ABZ24689-709),  
CC and to the use of these for the diagnosis, treatment or prevention  
CC of cell proliferative disorders including cancer, developmental  
CC disorders, neurological disorders, autoimmune disorders,  
CC reproductive disorders, and disorders of the placenta, and in the  
CC assessment of the effects of exogenous compounds on the activity  
CC and expression of proteins and nucleic acids associated with cell  
CC growth, differentiation and death. CGDD polynucleotides are also  
CC used in a claimed microarray and in a claimed method of generating  
CC an expression profile of a sample.  
XX Sequence 7742 BP; 2388 A; 1423 C; 1622 G; 2309 T; 0 other;  
SQ  
Query Match 98.3%; Score 5115; DB 25; Length 7742;  
Best Local Similarity 98.8%; Pred. No. 0;  
Matches 5168; Conservative 10; Mismatches 6; Indels 45; Gaps 1;  
QY 22 GGTACTGAGAGATGGAATACAGCGGAGTTACCCAGACCCCTCAGCGCTCGGCATCT 81  
DB 1 GGTACTGAGAGATGGAATACAGCGGAGTTACCCAGACCCCTCAGCGCTCGGCATCT 60  
QY 82 TGGTGGGATCAGCAAGTTCATTTTATCTGCTTTCTTCATCATTTTGGCACAATTTGGTG 141  
DB 61 TGGTGGGATCAGCAAGTTCATTTTATCTGCTTTCTTCATCATTTTGGCACAATTTGGTG 120  
QY 142 CCAGAAATTTACTTTGCTGAAATGGACCCAGACTTGGAAAAGCAGGAGGAAAGTGTACAA 201  
DB 121 CCAGAAATTTACTTTGCTGAAATGGACCCAGACTTGGAAAAGCAGGAGGAAAGTGTACAA 180  
QY 202 ATGTCAATATTCACCTCCACTGGATGGTACTTATTTGGAGAGATCCAGATATTTGCTTA 261  
DB 181 ATGTCAATATTCACCTCCACTGGATGGTACTTATTTGGAGAGATCCAGATATTTGCTTA 240

QY 262 GAGAAATTGAAGCACAGTGGAGCATTTACGCTTTGTGGGAGGGTTTTCAAAAGTGGAGAG 321  
DB 241 GAGAAATTGAAGCACAGTGGAGCATTTACGCTTTGTGGGAGGGTTTTCAAAAGTGGAGAG 300  
QY 322 ACAACCTATTCTTGCAGGGATTGTGCAATTTGATCCACACATGTGTACTCTGTATGGACTGC 381  
DB 301 ACAACCTATTCTTGCAGGGATTGTGCAATTTGATCCACACATGTGTACTCTGTATGGACTGC 360  
QY 382 TTCAGGACAGTGTTCATAAAATTCATCGTTTACAAGATGCATCTTACTTGGAGGAGGG 441  
DB 361 TTCAGGACAGTGTTCATAAAATTCATCGTTTACAAGATGCATCTTACTTGGAGGAGGG 420  
QY 442 TTCTGTGACTGTGGAGACACAGAGGCATGAAAACATGCGCCCTTTTGTGTAAATCATGAA 501  
DB 421 TTCTGTGACTGTGGAGACACAGAGGCATGAAAACATGCGCCCTTTTGTGTAAATCATGAA 480  
QY 502 CTTGGAAGACAGTACTATAAAGAGAAATTCACGCTGTCGTTGTAATGAAGAGGTAATT 561  
DB 481 CTTGGAAGACAGTACTATAAAGAGAAATTCACGCTGTCGTTGTAATGAAGAGGTAATT 540  
QY 562 GTCCAAAGCCAGGAAATATTTCCCTTCAGTGATAAAATATGTCGTAGAAATGACTATATGG 621  
DB 541 GTCCAAAGCCAGGAAATATTTCCCTTCAGTGATAAAATATGTCGTAGAAATGACTATATGG 600  
QY 622 GAAGAGAAAAGAACTGCTCTGAACTCCAGATAAGGKNR----- 663  
DB 601 GAAGAGAAAAGAACTGCTCTGAACTCCAGATAAGGGAGAAAAATGAAAGATACTAT 660  
QY 664 -----YVCVNDHSHSYDHTGCATATACAGCCTACAAGA 696  
DB 661 TGTGCTCTTTTCAATGATGAACACCATTCATATACCCAGTCATATACAGCCTACAAGA 720  
QY 697 GCTCTTGACTGTGAGCTGCGAGAGCCAGTTGTCATACCATCGCCATTTGACAAAAGAGGT 756  
DB 721 GCTCTTGACTGTGAGCTGCGAGAGCCAGTTGTCATACCATCGCCATTTGACAAAAGAGGT 780  
QY 757 CGTCGGGCTTTAAAGCGGGAGCTTATGCTGCTTGCAGGAGCAAGGAAAGATATAAG 816  
DB 781 CGTCGGGCTTTAAAGCGGGAGCTTATGCTGCTTGCAGGAGCAAGGAAAGATATAAG 840  
QY 817 AGTCATTCAAAAATGCTCTCAACATCCACTTTCATGTAGAAATTTACACTCAGAGATT 876  
DB 841 AGTCATTCAAAAATGCTCTCAACATCCACTTTCATGTAGAAATTTACACTCAGAGATT 900  
QY 877 ATGCTCATCAGAAATTTGCTTTTCGCTCTTGGTTTCCCTGGATGAACAAAATTTATGAGCTAT 936  
DB 901 ATGCTCATCAGAAATTTGCTTTTCGCTCTTGGTTTCCCTGGATGAACAAAATTTATGAGCTAT 960  
QY 937 TCAAGTGACTTTAGGACAGATCTTTTGGCAAGCATGCTTTAGAGAGAACCTGACTCGGAG 996  
DB 961 TCAAGTGACTTTAGGACAGATCTTTTGGCAAGCATGCTTTAGAGAGAACCTGACTCGGAG 1020  
QY 997 AATCCCTGCTCTATAAGCAGGTTAATGCTTTGGGATGCAAAAGCTTTATAAAGGTGCCCGT 1056  
DB 1021 AATCCCTGCTCTATAAGCAGGTTAATGCTTTGGGATGCAAAAGCTTTATAAAGGTGCCCGT 1080  
QY 1057 AAGATCCTTCATGAATTTGATCTTCAGCAGATTTTATTTATGGAGATGGAATACAAAACATC 1116  
DB 1081 AAGATCCTTCATGAATTTGATCTTCAGCAGATTTTATTTATGGAGATGGAATACAAAACATC 1140  
QY 1117 TTTGCTATGGAATTTGTGAAGTATTATAAACAACTGCAGAAAGAAATATATCAGTGATAT 1176  
DB 1141 TTTGCTATGGAATTTGTGAAGTATTATAAACAACTGCAGAAAGAAATATATCAGTGATAT 1200  
QY 1177 CATGACAGAGATATCTCTATAAATGCACTTTTCAGTTTCAGATGTTTACTGTTCTCTACTCTG 1236  
DB 1201 CATGACAGAGATATCTCTATAAATGCACTTTTCAGTTTCAGATGTTTACTGTTCTCTACTCTG 1260  
QY 1237 GCTCGACATCTTATGAAGACAGCAATGTTATCTGTCATTTACTTACTGAAACTCTGCTAGAA 1296  
DB 1261 GCTCGACATCTTATGAAGACAGCAATGTTATCTGTCATTTACTTACTGAAACTCTGCTAGAA 1320

OY 1297 GTTTACCTGAGTACTGGACAGAACATAATTCACTCCAGGGTTATAGCCGAGAC 1356  
 |||||  
 Db 1321 GTTTACTGAGTACTGGACAGAACATAATTCACTCCAGGGTTATAGCCGAGAC 1380  
 OY 1357 AAATTGGGAAGATATATGACCTAATATGTGACCTAAATATATCTGATCGCAAAACC 1416  
 |||||  
 Db 1381 AAATTGGGAAGATATATGACCTAATATGTGACCTAAATATATCTGATCGCAAAACC 1440  
 OY 1417 ACAATATGACAGAAAGATTAAGAAATGCAAGTTCCTGAAGTTTCGATCTTTTGAAG 1476  
 |||||  
 Db 1441 ACAATATGACAGAAAGATTAAGAAATGCAAGTTCCTGAAGTTTCGATCTTTTGAAG 1500  
 OY 1477 ATCTTACTGATATGACAGGAATGGAAGAAATCCGAGACAGTGGGCAACAATGAA 1536  
 |||||  
 Db 1501 ATCTTACTGATATGACAGGAATGGAAGAAATCCGAGACAGTGGGCAACAATGAA 1560  
 OY 1537 GTGATCTGATTTGGGAGGCTGCCATGTCTATACAGATGCAATGAGAAATTTTACAC 1596  
 |||||  
 Db 1561 GTGATCTGATTTGGGAGGCTGCCATGTCTATACAGATGCAATGAGAAATTTTACAC 1620  
 OY 1597 ATGTTCCAGAGTGGTGTCTGTGTGATGAGAACTCTTACTTGTGGCTTATAAAGATGT 1656  
 |||||  
 Db 1621 ATGTTCCAGAGTGGTGTGTGTGATGAGAACTCTTACTTGTGGCTTATAAAGATGT 1680  
 OY 1657 CACAAAGCTGTGATGAGGTGCAGTACCAAGTTTCATATCTAGTAGCAAGACAGTAGTACA 1716  
 |||||  
 Db 1681 CACAAAGCTGTGATGAGGTGCAGTACCAAGTTTCATATCTAGTAGCAAGACAGTAGTACA 1740  
 OY 1717 TCGTGTGACATAGTTTGGAAACAAGTCTTACAGATATCTGAGATCTTGTAAAGCATA 1776  
 |||||  
 Db 1741 TCGTGTGACATAGTTTGGAAACAAGTCTTACAGATATCTGAGATCTTGTAAAGCATA 1800  
 OY 1777 CATCGCACACTCTAGAGACCTTGTGTGTCTCATGTACGTTTAAAGCAGGCTGGTGT 1836  
 |||||  
 Db 1801 CATCGCACACTCTAGAGACCTTGTGTGTCTCATGTACGTTTAAAGCAGGCTGGTGT 1860  
 OY 1837 GTTTCAGACATGACGAATTTGTCTTTTGAAGACTTTCAGATGAGAGTACTAGTGA 1896  
 |||||  
 Db 1861 GTTTCAGACATGACGAATTTGTCTTTTGAAGACTTTCAGATGAGAGTACTAGTGA 1920  
 OY 1887 TATCTTTACGTTGTCTGTGTGTGTGCCAGGTTGTGCTGAGATGTGGCGAAGAAAT 1956  
 |||||  
 Db 1921 TATCTTTACGTTGTCTGTGTGTGTGCCAGGTTGTGCTGAGATGTGGCGAAGAAAT 1980  
 OY 1957 GGACGTCTCTTATTAAGCAGGCTGTTTATTAACAAGATGTAATGACAGAAAGAAAG 2016  
 |||||  
 Db 1981 GGACGTCTCTTATTAAGCAGGCTGTTTATTAACAAGATGTAATGACAGAAAGAAAG 2040  
 OY 2017 TATGATTAAGATATCATCATGCTTCAGATTTGGTGCATCTTTATGATGATCCCAATTAAGTTC 2076  
 |||||  
 Db 2041 TATGATTAAGATATCATCATGCTTCAGATTTGGTGCATCTTTATGATGATCCCAATTAAGTTC 2100  
 OY 2077 TTGTTACTGTAATTCAGAGGTATGAACTTGCCGAGGCTTTTAAAGACCATATCTACA 2136  
 |||||  
 Db 2101 TTGTTACTGTAATTCAGAGGTATGAACTTGCCGAGGCTTTTAAAGACCATATCTACA 2160  
 OY 2137 AAAAGACCGAGATTGATTAACAAATATATACACTTAATAGAAGAAATGCTTAGAGTCTTC 2196  
 |||||  
 Db 2161 AAAAGACCGAGATTGATTAACAAATATATACACTTAATAGAAGAAATGCTTAGAGTCTTC 2220  
 OY 2197 ATCTATATTTGGGAGCGTATATGTAAGTGGGAAATGTACCCAAAAGAGGCTC 2256  
 |||||  
 Db 2221 ATCTATATTTGGGAGCGTATATGTAAGTGGGAAATGTACCCAAAAGAGGCTC 2280  
 OY 2257 ACAATGAGAGAAATCATTCACCTTGTTCATTTGAACCATGCCACAGTGCCTTATGCC 2316  
 |||||  
 Db 2281 ACAATGAGAGAAATCATTCACCTTGTTCATTTGAACCATGCCACAGTGCCTTATGCC 2340  
 OY 2317 AAAAATTTTACCTGAGATGAATAATATGAACCTGGCTTAGAGAAATGTCTATAAACAAGTG 2376  
 |||||  
 Db 2341 AAAAATTTTACCTGAGATGAATAATATGAACCTGGCTTAGAGAAATGTCTATAAACAAGTG 2400  
 OY 2377 GCCACATTTAAGAAACGAGTGTATCAGGCCATGAGATTATGAACATAAAGATGATCA 2436

Db 2401 GCCACATTTAAGAAACGAGTGTATCAGGCCATGAGTTTATGAACATAAAGATGAATCA 2460  
 |||||  
 OY 2437 CTGAAGACTCTCAATATGATCTTTATCATCTACTCCAAAACCCAGCATAGCAAGGCTGAA 2496  
 |||||  
 Db 2461 CTGAAGACTCTCAATATGATCTTTATCATCTACTCCAAAACCCAGCATAGCAAGGCTGAA 2520  
 OY 2497 CATATGAGAGAAAGAGAGAAACAAGAAAAAAGATGAGCAATTTGCCCCACACCA 2556  
 |||||  
 Db 2521 CATATGAGAGAAAGAGAGAAACAAGAAAAAAGATGAGCAATTTGCCCCACACCA 2580  
 OY 2557 CCTCTGATTTCTGCCCTGCTTTTACGAAAGTGTATTAACCTTCTCACTGTGATATCATG 2616  
 |||||  
 Db 2581 CCTCTGATTTCTGCCCTGCTTTTACGAAAGTGTATTAACCTTCTCACTGTGATATCATG 2640  
 OY 2617 ATGTACATCTCAGGACCGTATTTGAGCGGGCATATGACAGAAATTCATCTGTGGACC 2676  
 |||||  
 Db 2641 ATGTACATCTCAGGACCGTATTTGAGCGGGCATATGACAGAAATTCATCTGTGGACC 2700  
 OY 2677 GAAGGATGCTCCAAATGCGCTTTTCAATTTCTGACATTTGGGTTTACTAGAGAGACAA 2736  
 |||||  
 Db 2701 GAAGGATGCTCCAAATGCGCTTTTCAATTTCTGACATTTGGGTTTACTAGAGAGACAA 2760  
 OY 2737 CAGCTTCAAAAGCTCCTGAGAGAAAGATTAACATTTGACATTTAATAGGCTTCAAGA 2796  
 |||||  
 Db 2761 CAGCTTCAAAAGCTCCTGAGAGAAAGATTAACATTTGACATTTAATAGGCTTCAAGA 2820  
 OY 2797 TTGGGAAGTTCAACCATGAATATACAAATGCTTTGGAAAAACCAAAAGAAATTCACAG 2856  
 |||||  
 Db 2821 TTGGGAAGTTCAACCATGAATATACAAATGCTTTGGAAAAACCAAAAGAAATTCACAG 2880  
 OY 2857 TTGAAGAGCCAGAGGACATGATTAACGTGATACTTACAGATGTTTGAACAGTAAAGCA 2916  
 |||||  
 Db 2881 TTGAAGAGCCAGAGGACATGATTAACGTGATACTTACAGATGTTTGAACAGTAAAGCA 2940  
 OY 2917 TTGAAGAAAAAATCTTGTTTAATTTGTAACAACCATCAGAGTGCATTTAATGAAT 2976  
 |||||  
 Db 2941 TTGAAGAAAAAATCTTGTTTAATTTGTAACAACCATCAGAGTGCATTTAATGAAT 3000  
 OY 2977 GATGACATTTACTATGATTAAGAAAAAGCAAGCAAGAAAAAGAAAGCTGAAGTGTGAG 3036  
 |||||  
 Db 3001 GATGACATTTACTATGATTAAGAAAAAGCAAGCAAGAAAAAGAAAGCTGAAGTGTGAG 3060  
 OY 3037 CTACATGCGCAGAGATCATGCGTCAGATGTCGCTTACAGAAAACTTATGAAT 3096  
 |||||  
 Db 3061 CTACATGCGCAGAGATCATGCGTCAGATGTCGCTTACAGAAAACTTATGAAT 3120  
 OY 3097 CATTAACATGATATGACATTAATCATGAGAAATGCTGGAAAGAAATTCATTAATGAG 3156  
 |||||  
 Db 3121 CATTAACATGATATGACATTAATCATGAGAAATGCTGGAAAGAAATTCATTAATGAG 3180  
 OY 3157 GAAGAGAGCAACCCAGCAGTCACTGATCTAGAAATGCTTTGGTCTTAAAGGGGT 3216  
 |||||  
 Db 3181 GAAGAGAGCAACCCAGCAGTCACTGATCTAGAAATGCTTTGGTCTTAAAGGGGT 3240  
 OY 3217 CCATCTGTTACTGAAAAAGAGGTGCTGACGTGATCTTTGCCAAGAAAGCAAGAGGTG 3276  
 |||||  
 Db 3241 CCATCTGTTACTGAAAAAGAGGTGCTGACGTGATCTTTGCCAAGAAAGCAAGAGGTG 3300  
 OY 3277 AAAATGAAAAATATCCATGATATATGCGGCTGTCTCAGAAATCTACAGCTTTAAC 3336  
 |||||  
 Db 3301 AAAATGAAAAATATCCATGATATATGCGGCTGTCTCAGAAATCTACAGCTTTAAC 3360  
 OY 3337 CAGCAGAGGGAAAAACCCATAGAACTCTCAGAGAAAGCCCTAGACCCACTTTTATGAGAT 3396  
 |||||  
 Db 3361 CAGCAGAGGGAAAAACCCATAGAACTCTCAGAGAAAGCCCTAGACCCACTTTTATGAGAT 3420  
 OY 3397 CCAGACTTGGCATATGGAATTTATACAGGAAGCTGTGCTATGTAATGACAGCAGTGTGC 3456  
 |||||  
 Db 3421 CCAGACTTGGCATATGGAATTTATACAGGAAGCTGTGCTATGTAATGACAGCAGTGTGC 3480  
 OY 3457 TGGCACAAATATTTTGAAGCTGTACAGCTAGAGCTCAGCAGCCATTCATGTTGACTT 3516  
 |||||

Db 3481 TGGCAGAAGTATTTTGAAGCTGTACAGCTCAGCTCTCAGCAGCGCATTCATGTTGACCTT 3540  
QY 3517 TTGTAGCTTGAAGTGGGAATATCTTTGCCCTCTTTGCAAAATCTCTGTGCAATACTGTG 3576  
Db 3541 TTGTAGCTTGAAGTGGGAATATCTTTGCCCTCTTTGCAAAATCTCTGTGCAATACTGTG 3600  
QY 3577 ATCCCATATTCTTTGCAACCTCAAAGATAAACAGTAGAGATGCAGATGCTCTTGCT 3636  
Db 3601 ATCCCATATTCTTTGCAACCTCAAAGATAAACAGTAGAGATGCAGATGCTCTTGCT 3660  
QY 3637 CAACTTTTGACCTGGCAGGTGGATACAGACTGTTCTTGCCCAAGATATCAGGTATAAT 3696  
Db 3661 CAACTTTTGACCTGGCAGGTGGATACAGACTGTTCTTGCCCAAGATATCAGGTATAAT 3720  
QY 3697 ATAAGACATGCTAAAGGAGAAACCCAAATTCCTATTCTTTTAATCAAGGAATGGAGAT 3756  
Db 3721 ATAAGACATGCTAAAGGAGAAACCCAAATTCCTATTCTTTTAATCAAGGAATGGAGAT 3780  
QY 3757 TCTACTTTGGAGTTCATTCCTAGTTCCTGAGTTTGGGGTGGAGTTCGATTAATATTC 3816  
Db 3781 TCTACTTTGGAGTTCATTCCTAGTTCCTGAGTTTGGGGTGGAGTTCGATTAATATTC 3840  
QY 3817 AATAGCATCAAGGAATGGTATTCTTTGCCACAACAATTTATAGAATTGGATTGAAA 3876  
Db 3841 AATAGCATCAAGGAATGGTATTCTTTGCCACAACAATTTATAGAATTGGATTGAAA 3900  
QY 3877 GTGCCACCTGATGAAGGAGTCTCGAGTCCCATGCTGACCTGGAGCACCTCGCTTTC 3936  
Db 3901 GTGCCACCTGATGAAGGAGTCTCGAGTCCCATGCTGACCTGGAGCACCTCGCTTTC 3960  
QY 3937 ACTATCCAGCAATGAAATCTATTGGGAGATGAAGGAAACCTCTGTTTGAGCACCT 3996  
Db 3961 ACTATCCAGCAATGAAATCTATTGGGAGATGAAGGAAACCTCTGTTTGAGCACCT 4020  
QY 3997 CAAATAGCAGCATATGCTGGAAGCATTAATGCAGTTTCAGTTGCACAGAGGATT 4056  
Db 4021 CAAATAGCAGCATATGCTGGAAGCATTAATGCAGTTTCAGTTGCACAGAGGATT 4080  
QY 4057 ACCTGCTCCTCAGTCTGATACAGAAACATCTGGTTCGCTCTCTATCAGTTGTTCTTCCT 4116  
Db 4081 ACCTGCTCCTCAGTCTGATACAGAAACATCTGGTTCGCTCTCTATCAGTTGTTCTTCCT 4140  
QY 4117 AACATAAATCAGAAGATACACCATGCTCTCTCTATAGATCTGTTTCATGTTTGGTG 4176  
Db 4141 AACATAAATCAGAAGATACACCATGCTCTCTCTATAGATCTGTTTCATGTTTGGTG 4200  
QY 4177 GGTGCTGTGTAGCATTCCTGATGTTGGATGACCTCTGTTGATCTGCAGCCTTCT 4236  
Db 4201 GGTGCTGTGTAGCATTCCTGATGTTGGATGACCTCTGTTGATCTGCAGCCTTCT 4260  
QY 4237 TCAGTTAGTCTTCCCTATACCACTTTATCTCTCCATTTGATCACCATTGGCACACATG 4296  
Db 4261 TCAGTTAGTCTTCCCTATACCACTTTATCTCTCCATTTGATCACCATTGGCACACATG 4320  
QY 4297 CTTTCAGATACTACTACAGTAGACAGGCTTACCCTTGCTCAGGTTCAAGAGACAGT 4356  
Db 4321 CTTTCAGATACTACTACAGTAGACAGGCTTACCCTTGCTCAGGTTCAAGAGACAGT 4380  
QY 4357 GAAGAGGCTCATTCGCTCTCTCTTCTTGGAGAAATTTCTCAATATACAGTGGCTC 4416  
Db 4381 GAAGAGGCTCATTCGCTCTCTCTTCTTGGAGAAATTTCTCAATATACAGTGGCTC 4440  
QY 4417 ATTGGGTGTGATATTCCTGGCTGGTATTTGTTGGGTCTCACTGAAGATGCATCACCCT 4476  
Db 4441 ATTGGGTGTGATATTCCTGGCTGGTATTTGTTGGGTCTCACTGAAGATGCATCACCCT 4500  
QY 4477 TATCTTCGCTGTGCTGCATTTGTTTCCACTATTTACTTGGGTAACTCCGCTGAGGAA 4536  
Db 4501 TATCTTCGCTGTGCTGCATTTGTTTCCACTATTTACTTGGGTAACTCCGCTGAGGAA 4560  
QY 4537 CTGCATACCAATCTCCAGAGGAGATACAGTGCACCTCTGTAGCTATCTATTACCT 4596  
Db 4561 CTGCATACCAATCTCCAGAGGAGATACAGTGCACCTCTGTAGCTATCTATTACCT 4620

QY 4597 ACAAAATTTGTTCTGCTCTTCCAGAAATATTGGGATACTGTAAGGCCCTTGTCTCAGAGG 4656  
Db 4621 ACAAAATTTGTTCTGCTCTTCCAGAAATATTGGGATACTGTAAGGCCCTTGTCTCAGAGG 4680  
QY 4657 CGGTGTGCAGATCTGCTCTTAACTGTTTGAAGCAAAACACCGTGGTCAAGTAC 4716  
Db 4681 TGGTGTGCAGATCTGCTCTTAACTGTTTGAAGCAAAACACCGTGGTCAAGTAC 4740  
QY 4717 CCTAGAAAAGAAATAGTTTGTATAGAGCTTCTCTGATGACTATAGTGTCTCTGAATCAA 4776  
Db 4741 CCTAGAAAAGAAATAGTTTGTATAGAGCTTCTCTGATGACTATAGTGTCTCTGAATCAA 4800  
QY 4777 GCTTCTCATTTTCAGGTGCCACGGTCTGCAGATGATGAGCGAAAGCATCTCTCTCTGC 4836  
Db 4801 GCTTCTCATTTTCAGGTGCCACGGTCTGCAGATGATGAGCGAAAGCATCTCTCTCTGC 4860  
QY 4837 CTTTTCTGTGGGGTATACTATGTTCTCAGAACATTTTCTGCCAGGAAATTTGTGAACGG 4896  
Db 4861 CTTTTCTGTGGGGTATACTATGTTCTCAGAACATTTTCTGCCAGGAAATTTGTGAACGG 4920  
QY 4897 GAAGAGTTGGAGCTTGCAATTTTTCACGCACTTCACTGTGAGCGGAGTCTGCAATTTTC 4956  
Db 4921 GAAGAGTTGGAGCTTGCAATTTTTCACGCACTTCACTGTGAGCGGAGTCTGCAATTTTC 4980  
QY 4957 CTAAAAATCAGAGAAATGCCAGTGGTCTGTGTTGAAGTAAAGCCAGAGGCTGTGCCCTAT 5016  
Db 4981 CTAAAAATCAGAGAAATGCCAGTGGTCTGTGTTGAAGTAAAGCCAGAGGCTGTGCCCTAT 5040  
QY 5017 CCAGCTCCTTACTTGGATGAATATGGAGAAACAGACCTTGGCTGAAGGGGGCAACCCC 5076  
Db 5041 CCAGCTCCTTACTTGGATGAATATGGAGAAACAGACCTTGGCTGAAGGGGGCAACCCC 5100  
QY 5077 CTTTCAATTTATCTGCTGAGCGGTATCGGAAGTCCATTTTGGTCTGGCAACAACACTGCAATT 5136  
Db 5101 CTTTCAATTTATCTGCTGAGCGGTATCGGAAGTCCATTTTGGTCTGGCAACAACACTGCAATT 5160  
QY 5137 ATAGAAGAGATTTGTAGAGCCAAAGAGACTAATCAGATGTTTATTTGGATTCAACTGGCAG 5196  
Db 5161 ATAGAAGAGATTTGTAGAGCCAAAGAGACTAATCAGATGTTTATTTGGATTCAACTGGCAG 5220  
QY 5197 TTACTGTGA 5205  
Db 5221 TTACTGTGA 5229  
RESULT 2  
ID AAV99308 standard; cdna; 6395 BP.  
XX AAV99308;  
XX AC AAV99308;  
XX XX 25-MAR-1999 (first entry)  
XX DE cdna encoding mouse a ubiquitin-protein ligase, Ubql.  
XX KW Ubiquitin-protein ligase; Ubql; mouse; ubiquitinylation; degradation;  
KW K end rule pathway; stress-related muscle wasting; inhibitor; screen; ss.  
XX OS Mus sp.  
XX FH Key Location/Qualifiers  
FT CDS 115..5388  
FT FT /\*tag= a  
FT FT /product= Ubql  
XX US5861312-A.  
PN 19-JAN-1999.  
XX PD 02-DEC-1997; 97US-0982956.  
PF 02-DEC-1997; 97US-0982956.  
PR 02-DEC-1997; 97US-0982956.



XX (CALX) CALIFORNIA INST OF TECHNOLOGY.  
XX  
XX Kwon YT, Varshavsky A;  
XX  
XX WPI, 1999-130395/11.  
XX DR P-PSDB; AAM84351.  
XX  
XX Mouse and human Ubr1 cDNA - useful for producing recombinant Ubr1  
XX PT polypeptides  
XX  
XX Claim 1: Columns 15-28; 18pp; English.  
XX  
XX The present sequence encodes a ubiquitin-protein ligase called Ubr1.  
XX CC The Ubr1 enzymes are involved in protein ubiquitinylation and  
XX CC ultimate degradation through the N-end rule pathway and have been  
XX CC linked to stress-related muscle wasting. Recombinant Ubr1 polypeptides  
XX CC can be used to screen for inhibitors of muscle wasting when this is  
XX CC associated with the N-end rule pathway.  
SQ Sequence 6395 BP; 1802 A; 1376 C; 1519 G; 1698 T; 0 other;  
  
Query Match 78.5%; Score 4085.6; DB 20; Length 6395;  
Best Local Similarity 86.6%; Pred. No. 0;  
Matches 4567; Conservative 9; Mismatches 629; Indels 69; Gaps 4;  
  
1 ATGCGGACGAGAGGCTGGAGTACTGAGAGAGTGAATCAGCCGGAGTTACCCAG 60  
115 ATGGCGGACGAGAGTGGACGCGCCGAGAGATGAGCTGACGCCGGAGCCTCCCTG 174  
  
61 ACCCTCAGCGTGGCATCTTGGTGGATGATCAGCAAGTGGATTTTACTGCTTCTG 120  
175 GCCCGGACGCGCGGCTGCTGTTGGATGAGCAAGTGGATTTCTATCTGCTTCTTA 234  
  
121 CATCATTTGGCAATTTGGTGGCAGAAATTTACTTTGCTGAATGGAGCCAGATTTGAA 180  
235 CATCATTTGGCAGATTTAGTGGCAGAAATTTATTTTGGTGGATGGACCCAGATTTGAA 294  
  
181 AAGCAGAGAGAAAGTGCATCAATATTTCACTCCACTGAGATGATTTATTTTGA 240  
295 AAGCAGAGAGAGATGATCAGATTCATCACTCTTGGAGTGGATTTATTTTGA 354  
  
241 GAAGATCCAGATTTTGTCTTGAAGAAATTTGAAGCAGATTTAGCTTTTGG 300  
355 GAGGATCCGATATTTGCTTGAAGAAATTTAAACAGTGGAGCGTTCCAGTTGTGG 414  
  
301 AAGGATCCAAAGAGAGAGACACCTATTTCTTGGAGGATTTGCAATTTGCAACA 360  
415 AAGGATTTCAAAAGTGGAGAAACATATTTCTTGGAGATTTGCAATTTGCAACA 474  
  
361 TGTGTACTCTGTATGAGTCTTCCAGACAGTGTTCATAAATCATCGTTACAGATG 420  
475 TGTGTCTCTGTATGAGTCTTCCAAAGTATGTTCAATAAACATCGTTACAGATG 534  
  
421 CATACTTCTACTGAGAGAGGTTTGTGTGACTGTGAGACACAGAGGATGAAACTGGC 480  
535 CATACTTCTACTGAGAGGTTTGTGTGACTGTGAGACACAGAGGATGAAACTGGC 594  
  
481 CTTTGTGTAAATCATGAAACCTGGAAGAGAGAGTACTATAAAGAGATTCACCTGT 540  
595 CTTTGTGTGTGATCAGACGCTGGAAGAGAGTACTATAAAGAGAGTACTATTC 654  
  
541 CCGTGAATGAAGAGTATTTGTCAGAGCAGAGAAATTTCTTCACTGATTAATATAT 600  
655 CCATTGAATGAAGAGTATTTGTCAGAGCAGAGATATTTCTTCCGTTGATTAATAT 714  
  
601 GTCTGAATGATCTATATGAGAGAGAGAGAAAGAACTGCTCTGAACTCCAGATTAAG 660  
715 AATTGAGAAATGATCTATATGAGAGAGAGAGAAAGAAATGCTCTGAACTCCAGATTAAG 774  
  
661 KNR-----YYCVDNHHSDYH 675  
775 GAGAAATGAAGATCTATTTGTCTTTCAGAGATGAGACACATTCGATATATAT 834

QY 676 GTCATATACAGCTTACAAAGAGCTTTGATGCTGACGCTGAGAGGCCAGTTGCATACC 735  
DB 835 GTGATCTACAGTCTGACAGAGAGCTTGAATTTGCGAGCTTGGACAGACAGCTGCACAG 894  
QY 736 ACTGCCATTTGCAAGAGAGGCTGCGGCTGTTAAAGGGGAGCTTATGCTGCTTCCAG 795  
DB 895 ACTGCCATTTGCAAGAGAGGCTGCGGCTGTTCAAGAGAGGCTTATGCTGCTTCCAG 954  
QY 796 GAAGCAAGAGAGATATTAAGAGTCAATTCAGAAATGCTCTCAACATCCACTTCATGTA 855  
DB 955 GAAGCAAGAGAGATATTAAGAGTCAATTCAGAAAGAGTCTCTCACACACCCCTCATATG 1014  
QY 856 GAAGTATTAACCTGACAGATTTATGCTCATAGAAATTTGCTTCCGCTTGTCTG 915  
DB 1015 GAAGTGTGACCTGCTGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1074  
QY 916 ATGACCAAAATTTATGACTTATTCAGAGACTTTGAGGAGATTTTGGCAAGGCTT 975  
DB 1075 ATGACCAAAATTTATGACTTATTCAGAGACTTTGAGGAGATTTTGGCAAGGCTT 1134  
QY 976 AGAGAAAGACCTGACCTGAGAGAAATCCCTGCTCATTAACAGAGTTAATGCTTGGATGA 1035  
DB 1135 GTAGAAAGACCTGCTGAGAAATCCCTGCTCATTAACAGAGTTAATGCTTGGATGA 1194  
QY 1036 AAGCTTTATTAAGTGTCCCGTAAAGATCCTTCATGATTTGATTCAGAGATTTTATG 1095  
DB 1195 AAGCTTTATTAAGTGTCCCGTAAAGATCCTTCATGATTTGATTTATG 1254  
QY 1096 GAGATGAATTAACAAACCTTGTGATGAAATTTGAGAGTATTAACAACTGAG 1155  
DB 1255 GAGATGAATTAACAAACCTTGTGATGAAATTTGAGAGTATTAACAACTGAG 1314  
QY 1156 AAGAAATATATAGTATGATCATAGACAGATATCTTATTAACAGACTTCACTGAG 1215  
DB 1315 AAGAGATATATAGTATGATCATAGACAGATATCTTATTAACAGACTTCACTGAG 1374  
QY 1216 ATGTTACTGTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1275  
DB 1375 ATGTTACTGTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1434  
QY 1276 ATTACTGAATCTGCTGAGAGATTTTACCTGATGAGTGGAGAGCAACAAATTAATCAAC 1335  
DB 1435 ATTACTGAATCTGCTGAGAGATTTTACCTGATGAGTGGAGAGCAACAAATTAATCAAC 1494  
QY 1336 TTCAGAGGTTATAGCCAGCAAAATTTGAGAGATATTAAGATTAAGATTAAGATTAAG 1395  
DB 1495 TTCAGAGGTTATAGCCAGCAAAATTTGAGAGATATTAAGATTAAGATTAAGATTAAG 1554  
QY 1396 TATATCTGATCAGCAAAATTTGAGAGATATTAAGATTAAGATTAAGATTAAGATTAAG 1455  
DB 1555 TATATCTGATCAGCAAAATTTGAGAGATATTAAGATTAAGATTAAGATTAAGATTAAG 1614  
QY 1456 GGTTCGATCTTTTGAAGATTTCTTACCTGATGAGGAGATTAAGATTAAGATTAAGATTAAG 1515  
DB 1615 GGTTCGATCTTTTGAAGATTTCTTACCTGATGAGGAGATTAAGATTAAGATTAAGATTAAG 1674  
QY 1516 CAGGTTGGAGACACATTTGAAGTATTCATTTGAGGAGGCTGCTGCTGCTGCTGCTGCTG 1575  
DB 1675 CAGGTTGGAGACACATTTGAAGTATTCATTTGAGGAGGCTGCTGCTGCTGCTGCTGCTG 1734  
QY 1576 CAATTAAGATATTTTACTCATTTTCAAGAGAGTGTGCTTGTGATGAAGAACTTTA 1635  
DB 1735 CAATTAAGATATTTTACTCATTTTCAAGAGAGTGTGCTTGTGATGAAGAACTTTA 1794  
QY 1636 CTGTGCTTATTAAGATGTCAGAAAGCTGTGATGAGTGCAGTACAGTTTCAATCT 1695  
DB 1795 CTGTGCTTATTAAGATGTCAGAAAGCTGTGATGAGTGCAGTACAGTTTCAATCT 1854  
QY 1696 AGTAGCAGACAGTATGATGATGCTGTGAGCATATGTTGGAACAAAGTCTTACAGATTA 1755  
DB 1855 AGTAGCAGACAGTATGATGATGCTGTGAGCATATGTTGGAACAAAGTCTTACAGATTA 1914

QY	1756	TCTGAGGATCTTGTAAAGCATACATCTGCCACTCTCTAGGACCCCTTCTGCTCTTCATGTA	1815	Db	2995	CTCTTGGAAAGACTCAAGGAATCCCCCAATTAGAGCCCAAGGACATGATACATGCG	3054
Db	1915	TCAGAGACCTTGTAAAGCATACCTGCCACTCTCTAGAACACTTCTGCTGCTTCATGTA	1974	QY	2887	ATACTTCAGATGTTTGACACAGTGAAGCGATTAAAGAGAAAAATCTTGTAAATGTAGCA	2946
QY	1816	CGTTTAAAGCAGGCTGGTCTGTTTCAAGACTGCATGAATTTGTGCTTTTGAGGACTTT	1875	Db	3055	ATACTCCAGATGTTTGACACAGTGAAGCGATTAAAGAGAAAAATCTTGTAAATGTAGCA	3114
Db	1975	CGTTTAAAGCAGACTAGTGCTATTTCAGACTGCATGAATTTGTGCTTTTGACAGCTTT	2034	QY	2947	ACCACATCAGATCGGAAATCTATTAAAGATGATGAGATTACTCATGATAAAGAAAAACA	3006
QY	1876	CAAGTAGAGGTACTAGTGAATATCCCTTTACGTTGTCTGGTGTGTCGCCAGGTTGTT	1935	Db	3115	ACCACCTTCAGGACTGGAGTGCATTTAAGATGAGGAGATTACTCATGATAAAGAAAAACA	3174
Db	2035	CAAGTAGAGGTCTGCTGGAGTACCCGCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	2094	QY	3007	GAACGAAAAAGAAAGCTGAAGCTGCTAGGCTACATGCCAGAGAGATCATGGCTCAGATG	3066
QY	1936	GCTGAGATCTGGGAGAAATGGACTGCTCTATTATAGCCAGTGTTTATTACCAAGAT	1995	Db	3175	GAACGGAAGAGAAAGCTGAGGCCCTGAGGCTTCTATGCCCAAGAGATCATGGCCAGATG	3234
Db	2095	GCTGAGATCTGGGAGAAAGCGGCTCTCACTCATCAGCCAGGTTTCTTATTATCAAGAT	2154	QY	3067	TCTGCCCTACAGAAAAAATCTTCAATTTGAACCTCATAACTCATGTATGACAATACATACAA	3126
QY	1996	GTTAAAGTCAGAGAAAGAAATGTATGATAAAGATATCATGCTTCAGATTTGTCATCT	2055	Db	3235	TCTGCCCTACAGAAAAAATCTTCAATTTGAACCTCATAACTCATGTATGATAATACGTCAGAA	3294
Db	2155	GTTAAATGCAGGAGAAATGTACGATAAAGATATCATGCTTCAGATTGGAGCATCT	2214	QY	3127	ATGCGCTGGGAAAGAGATTCATTTATGGAGGAAGAGACCCAGAGAGTCAAGTACATAC	3186
QY	2056	TTAATGGATCCCAATTAAGTCTTGTGTACTTCTGAGTTCAGAGGATGAACTTGCCGAGGT	2115	Db	3295	GTAACAGGGAAGGAAGACTCCATTTATGGAGGAAGAGACCTCAGCAGTCAAGTGAGGCC	3354
Db	2215	ATAATGGATCCCAACAAGTCTTGTGTACTTCTGAGATATGAACTTCACTGATGCT	2274	QY	3187	TCTAGAATTCCTTTGGGCTCTTAAACGGGTCCTACTGTTACTGAAAAAGAGTCTGAGC	3246
QY	2116	TTTAAACAAGACCATATCTACAAAAGACAGGATTTGATTTAAACAATATATACACTAATA	2175	Db	3355	TCTAGAATTCCTTTGGGCTCTTAAACGGGTCCTACTGTTACTGAAAAAGAGTCTGAGC	3414
Db	2275	TTTAAACAAGACCATATCTCACAAGAACAGGATTTGATTTAAACAGTATATACATTAATA	2334	QY	3247	TGCATCTTTCGCCAAGAAAGACAGGAGTGAATAATAGAAAAATATGCCATGTATATTCG	3306
QY	2176	GAGAAATCTCTCAGCTCCTCATCTATATTGTGGGTGAGGTTATGCTACCTGAGTGGGA	2235	Db	3415	TGCATCTCTGCCAAGAAAGACAGGAGTGAATAATAGAAAAATATGCCATGTATATTCG	3474
Db	2335	GAGAAATCTCTCAGCTCCTCATCTATATTGTGGGTGAGGTTATGCTACCTGAGTGGGA	2394	QY	3307	GCCTGTGTCCCAAAATCTACTGCTTAAACCCAGCACAGGGGAAAAACCCATAGAACTCTCA	3366
QY	2236	AATGTGACCAAAAGAGAGTCACAATGAGAGAAATCATTCACCTTGCTTTGCAATGAAACCC	2295	Db	3475	GCATGTGTGCAGAAATCCACGCCCTTAACCCAGCACAGAGGAGCCCTGTGGACCACCTA	3534
Db	2395	AATGTTTACCAGAGAGAGGTATAATGAGAGAGATTACTACCTTACTTGTGATGAGCCC	2454	QY	3367	GGAGAAAGCCCTAGACCCACTTTTTCATGATCCAGACTTGGCATATGGAACCTTATACAGGA	3426
QY	2296	ATGCCACAGTGCCATTTGCCAAAAATTTACTCGAAGATGAATAATGAAACTGGCTTA	2355	Db	3535	GGGAAACACTGGGACCTCTTTTCATGGATCCAGACTTGGCACAATGGAACCTTATACAGGA	3594
Db	2455	ATGCCACAGTGCCATCGCCAGAAACCTACCTGAGAACGAAAAATGAAACTGGCTTA	2514	QY	3427	AGCTGTGGTCTATGTAATGCAGCAGTGTGCTGGCAGAAAGTATTTTGAAGCTGTACAGCTG	3486
QY	2356	GAGAAATCTATAAACAAGTGCCACATTTAAGAAACCCAGGTCATCAGGCCATGAGTT	2415	Db	3595	AGCTGTGGTCTATGTAATGCAGCAGTGTGCTGGCAGAAAGTATTTTGAAGCTGTACAGCTG	3654
Db	2515	GAGAAATCTATAAACAAGTGCCACATTTAAGAAACCCAGGTCATCAGGCCATGAGTT	2574	QY	3487	AGCTCTCAGCAGCCATTCATGTTGACCTTTTGTGACTTTGAAAGTGGAGAAATCTTTGC	3546
QY	2416	TATGAATGAAAGATGAATCACTGAAAGACTTCAATATATGTACTTTTATCATCTTCCAAA	2475	Db	3655	AGCTCGCAGCAGCGCATTCAGCTAGACCTGTTGACCTGGAGAGCGGCGAGTACCTATGC	3714
Db	2575	TATGAATGAAAGATGAATCACTGAAAGACTTCAATATATGTACTTTTATCATCTTCCAAA	2634	QY	3547	CCTCTTTGCAAAATCTCTGTGCAATACCTGATCCCATTTATCTTTTCCAACTCAAAAG	3606
QY	2476	ACCAGCATAGCAAGGCTGAACATATGCAGAGAAAGGAGAAACAAGAAAAACAAGAT	2535	Db	3715	CCGCTCTGCAAGTCTCTCTGCAACACTGTCTATCCCATCATCCCTTTTCGAGCCGAGAG	3774
Db	2635	ACACAGCATAGCAAGGCTGAACATATGCAGAGAAAGGAGAAACAAGAAAAATAAAGAT	2694	QY	3607	ATAACAGTGAAGATGCAGATGCTTCTGCTCAACTTTTGAACCTGGCAGCGTGGATACAG	3666
QY	2536	GAAGCATTCGCCGCCACCCACCTCTCTGAATTCCTGCCCTGCTTTCAGCAAGTGAATTAAC	2595	Db	3775	ATCAACAGTGAAGATGCGGAGGCTCTTCTGCTCAACTTTTGAACCTGGCCCGGTGGATACAG	3834
Db	2695	GAAGCATTCGCCGCCACCCACCTCTCTGAAGATTCCTGCCCTGCTTTCAGCAAGTGAATTAAC	2754	QY	3667	ACTGTTCTGGCCAGAAATATCAGTTTATATATAGACATGCTTAAAGAGGAGAAACCC---CA	3723
QY	2596	CTTCTCAACTGTGATATCATGTATGATGATCTCTCAGGACCGCTATTGAGCGGGCAATAGAC	2655	Db	3835	ACTGCTCTGCGCAGAAATATCGGTTATATAATAAGCATGCTTAAAGAGGAGAAACCCAGCA	3894
Db	2755	CTGCTCAGCTGTGATGTTATGATATACATCTCAGGACCACTTTGAGCGGGCAGTGAC	2814	QY	3724	ATTCCTATTTTCTTTAATCAAGGAATGGAGATTCCTACTTTGGAGTTCCATTTCCATCTCTG	3783
QY	2656	ACAGATTCTAACTTGTGGACGGAAGGATGCTCCAAATGGCTTTTTCATATCTGGCATG	2715	Db	3895	GTTCTCTGTCTTGTGTTAATCAAGGAATGGGGATTCCACTTTTGTAGTTTTCATTTCCATCTCTG	3954
Db	2815	ACGAGCTCTAATCTGTGGACAGAGGATGCTGCAGATGCGTTCCCATATATTGGCACTG	2874	QY	3784	AGTTTTCGCGTGTAGTCTTCGATTAATAATTAATAATAGCATCAAGGAATGTTATCTCTC	3843
QY	2716	GGTTTACTAGAAGAGCAACAGCTTCAAAAAGTCTCCTGAAGAGAGTAAACATTTGAC	2775	Db	3955	AGTTTTCGAGTTTCTGCTTCGTTGAAATATTCAAAATAGTATCAAGGAATGTTATCTCTC	4014
Db	2875	GGCTTGTGGAAGAGAGCAGCAGCTTCAGAAAGCTCCTGAAGAGGAGTGGCTTTTTCGAC	2934	QY	3844	TTTGCCCAACAATTTATAGAAATGGATTGAAAGTCCCACTCTGATCAAGAGGATCTCTCA	3903
QY	2776	TTTTATCATAAAGCTTCAAGATTTGGGAAGTTTACGCCAT-----GAATATACAATG	2826	Db	4015	TTGCGCCACAACATTTACAGAAATGGCCCTGAAAGTGGCTCTGATGAAGTGAACACCCAGCA	4074
Db	2935	TTTTATCATAAAGCTTCAAGATTTGGGAAGTTTACGCCATGATGCTCAGAAATATACAATG	2994	QY	3904	GTCCCATGCTGACCTGGAGCAGCTGCGTTTCACTATCCAGGCAATTTGAAATCTATTG	3963
QY	2827	CTTTTGGAAAACTCAAGGAATTCGCCAGTTAGAGGCCAGAGGACATGATAACGTGG	2886				

4075 GTGCCATGATGACCTGAGACAGTGTGGCTTCACCATCCAGGCATTCGAAACCTGTTG 4134  
 QY 3964 GGAGATGAAGAAACCTCTGTTTGAGACATTCAAATAAGCAGATATGCTCGAAA 4023  
 Db 4135 GGAGATGAAGAAACCTCTATTTGGAGCATTCAAATAAGCAGATATGCTCGAAA 4194  
 QY 4024 GGAATTAATCAGTTTGCAGTGCAGAGAGATTACCTGTCAGAGTCCGATACAGAAA 4083  
 Db 4195 GCGCTAATGCACTTGGACATGCAAGAGGCTACCTGCGCTCAGGCTCCGATACACAAA 4254  
 QY 4084 CATCTGTCGTCTCTATCAGTGTCTTCTTCTTACATTAATCAGAAAGTACACCATCG 4143  
 Db 4255 CATCTGTCGTCT 4314  
 QY 4144 CTCTGTCATAGATCTGTTTATGTTTGGTGGTGTCTGTCTGTCTGTCTGTCTGTCT 4203  
 Db 4315 CTCTGTCGTCT 4374  
 QY 4204 TATTGGATGACCCGTTGATCTGACGCTCTGCTGATTAGTTCTCTATACCACTT 4263  
 Db 4375 TATTGGATGACACCGTGTGATCTGACGCTCTGCTGCTCTCTCTCTCTCTCTCTCT 4434  
 QY 4264 TATCTCTTCATTTGATGATGACATGACATGCTCTCTCTCTCTCTCTCTCTCTCTCT 4321  
 Db 4435 TATCTCTTCATGATGATGACATGACATGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4494  
 QY 4322 -----CAGGCTACCCCTTGTCTCTGATGTTCAAGAAAGACAGTAAAGGCTCTCT 4371  
 Db 4495 GATCTGTCTCAGGCGCCGCTGCTGATGAGGATGAGAGATGATGAGAGGCTCTCTCT 4554  
 QY 4372 GATCT 4431  
 Db 4555 GATCTGTCT 4614  
 QY 4432 CCTGGCTGATTTTGGGCT 4491  
 Db 4615 CCGGCTGTGATCTGCT 4674  
 QY 4492 GCAATGTTTTCACATTTTACTTGGGTAACCTCCCTCTCTCTCTCTCTCTCTCTCTCT 4551  
 Db 4675 GCACGTCTTTCACATTTTACTTGGGTAACCTCCCTCTCTCTCTCTCTCTCTCTCTCT 4734  
 QY 4552 GCAGAAAGAGATGACATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4611  
 Db 4735 GCTGAAGGAGATGACATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4794  
 QY 4612 CTTCTCCAGAAATTTGGATGATCTGTAAGCCCTCTCTCTCTCTCTCTCTCTCTCTCT 4671  
 Db 4795 CTTCTCCAGAAATTTGGATGATGATCTGTAAGCCCTCTCTCTCTCTCTCTCTCTCTCT 4854  
 QY 4672 GCGTACTTAACCTGTTTGAAGCAAAACACCGCTGTCTCTCTCTCTCTCTCTCTCTCTCT 4731  
 Db 4855 GCGTACTTAACCTGTTTGAAGCAAAACACCGCTGTCTCTCTCTCTCTCTCTCTCTCTCT 4914  
 QY 4733 AGTTGATAGAGCTTCT 4791  
 Db 4915 AGTTGATAGAGCTTCT 4974  
 QY 4792 TCGCCACGCT 4851  
 Db 4975 TCGCCACGCT 5034  
 QY 4852 ATACTATGTTTTCAGAAATTTGCTGCAAGAAATTTGTAAGGGGAGAGGTTGAGCT 4911  
 Db 5035 ATCTCTGTCT 5094  
 QY 4912 TCGATTTTTCACGATCT 4971  
 Db 5095 TCGATTTTTCACGATCT 5154  
 QY 4972 TCGCCAGTGTCT 5031  
 Db 5155 TCGAGGCTGTCT 5214

QY 5032 GATGAATATGAGAAACAGACCTTGGCTGAAAGAGGGGCAACCCCTCATTTATCTGCT 5091  
 Db 5215 GATGAATATGAGAAACAGACCTTGGCTGAAAGAGGGGCAACCCCTCATTTATCTGCT 5274  
 QY 5092 GAGCGGTATGGAAGCTTCATTTTGTCTGGCAACACCTGATATATGAGAGATTGCT 5151  
 Db 5275 GAGCGGTATGGAAGCTTCATTTTGTCTGGCAACACCTGATATATGAGAGATTGCT 5334  
 QY 5152 AGGAGCCAGAGACTAATCAGATGTTTATTTGATTCACTGGCAGTACTGTA 5205  
 Db 5335 CGGAGCCAGAGACTAATCAGATGTTTATTTGATTCACTGGCAGTACTGTA 5388

RESULT 3  
 AAC86933 standard; cDNA; 6395 BP.  
 AAC86933;  
 02-APR-2001 (first entry)  
 Nucleotide sequence of Mouse Ubr1 protein.  
 Ubr1: E3-type protein; ubiquitin system; ubiquitin-protein ligase;  
 N-end rule pathway; intracellular pathogen; Lysteria monocytogenes;  
 Yersinia enterocolitica; muscle wasting; infection; ss.  
 Mus sp.  
 FH Key Location/Qualifiers  
 FT CDS 115..538  
 FT /\*tag= a  
 FT /product= "Ubr1"  
 PN U6159732-A.  
 PD 12-DEC-2000.  
 PF 11-JAN-1999; 99US-0228317.  
 PR 02-DEC-1997; 97US-0982956.  
 PA (CALY ) CALIFORNIA INST OF TECHNOLOGY.  
 PI Kwon YT, Varshavsky A;  
 DR WPI: 2001-090278/10.  
 DR P-PSDB; AAB31162.  
 XX PT Inhibiting the N-end rule pathway in mammalian cells for treating  
 PT infections and various diseases associated with muscle tissue wasting,  
 PT by inhibiting the expression of ubrl gene  
 PS Claim 3; Column 15-28; 18pp; English.  
 CC The present sequence encodes a murine Ubr1 enzyme. Ubr1 is an E3-type  
 CC protein of the ubiquitin system. Specifically, it is a ubiquitin-protein  
 CC ligase. The enzyme is specific for destabilizing residues exposed at  
 CC the N-terminus of protein substrates. Inhibition of the expression of  
 CC Ubr1 gene in a cell results in inhibition of the N-end rule pathway.  
 CC The method is used for treatment of mammalian cells infected with an  
 CC intracellular pathogen, e.g. Lysteria monocytogenes or Yersinia  
 CC enterocolitica. Inhibition of N-end rule pathway is also useful for  
 CC treating various diseases associated with wasting of muscle tissue and  
 CC infections.  
 SQ Sequence 6395 BP; 1802 A; 1376 C; 1519 G; 1698 T; 0 other;  
 XX  
 Query Match 78.5%; Score 4085.6; DB 22; Length 6395;  
 Best Local Similarity 86.6%; Pred. No. 0;  
 Matches 4567; Conservative 9; Mismatches 629; Indels 69; Gaps 4;

QY	1	ATGGCGGACAGAGAGCGTGGAGGTACTGAGAGATGGAAATCAGCGCGGAGTTACCCAG	60
Db	115	ATGGCGGACGAAGAGATGGACGCGCGGAGAGATGGACGTCTACGCGCGGAGCCTCCCGT	174
QY	61	ACCCCTCAGCGTCTGCATCTTGGTGGGATCAGCAAGTTGATTTTATACGTCTTCTTG	120
Db	175	GCCCGGACGGCGCGCATCGTGGTGGATCAGCAAGTTGATTTTCTATACTGTTCTTTA	234
QY	121	CATCATTTGGCACAATTTGGTCCAGAAATTTACTTTGCTGAAATGGACCCAGACTTGGAA	180
Db	235	CATCATTTGGCACAATTTAGTGGCCAGAAATTTATTTGCTGAGATGACCCAGATTTGGAA	294
QY	181	AAGCAGAGGAAAGTGTACAAATGTCAAAATTCACCTCCACCTGGAATGGTACTTATTGGA	240
Db	295	AAGCAAGAAGAGAGTGTACAGATGTCAATACTCACCTCTTGGAGTGGTACTTATTGGA	354
QY	241	GAAGATCCAGATATTTCCTTAGAGAAATTTGAACACAGATGGAGCATTTACGTTTGTGGG	300
Db	355	GAGGATCCGGATATTTCCTTAGAGAAATTTAAACACAGATGGAGCGTTCCAGTTGTGTGG	414
QY	301	AGGGTTTTCAAAAAGTGGAGAGACAACCTATTCTTGCAGGGATTTGCAATTGATCAACA	360
Db	415	AAGTTTTTCAAAAGTGGAGAAACACATATCTCTAGGGATTTGCAATTGATCAACAG	474
QY	361	TGTGTACTCTGTATGACTGCTTCCAGGACAGTGTTCATAAAAATCATCTTTACAAGATG	420
Db	475	TGTGTCTCTGTATGACTGCTTCCAAAGTAGTGTTCATAAAAACCATCGTTACAAGATG	534
QY	421	CATACTTCTACTGGAGAGGGTTCGTGACTGTGGAGACACAGAGGGATGGAAAATCTGGC	480
Db	535	CATACTTCTACTGGAGAGGGCTTCTGTGACTGTGGAGACACAGAAGCGTGGAAAATCTGGC	594
QY	481	CCTTTTTGTCTAAATCATGAACCTTGAAGAGCAGGTACTATAAAGAGAAATTCACGCTGT	540
Db	595	CCTTTTTGTGTGATCACGAGCCTGGAGAGCAGGTACTACAAAGAGAGACTTACATTCG	654
QY	541	CCGTTGAATGAAGAGGTAAATTTGCCAAGCCAGGAAAAATTTCTCTTCAGTGTAAAAATAT	600
Db	655	CCATTGAATGAAGAGGTGATTGTCTCAAGCCAGGAGAAATTTCCCTTCGGTGTAAAAATAC	714
QY	601	GTGCTAGAAATGACTATATGGGAAGAGGAAAAAGAACTGCCTCCTGAACTCCAGATAAGG	660
Db	715	ATTGTAGAAATGACTATATGGGAAGAGGAAAAAGAAATTCCTCCTGAACTGCAGATAAGG	774
QY	661	KNR-----YVCVNDHSYDH	675
Db	775	GAGAAAAATCAACGATACTATTGTGCTCTTTTCAACGATGAGCACCATTCGTATGATCAT	834
QY	676	GTCAATACAGCCCTACAAAGAGCTCTTGACTGTGAGCTCGCAGAGGCCAGTTGTCATACC	735
Db	835	GTGATCTACAGCTGCAGAGAGCTCTAGATTGGAGCTTCAGAGGACACAGCTGCACACG	894
QY	736	ACTGCCATTGACAAAGAGGGTCTGGGCTGTGTTAAAGCGGGAGCTTATGCTGTGGCCAG	795
Db	895	ACTGCCATCACAAGAGGGTCCGCGGCTGTCAAAGCAGGTGTGTATGCCACTTGGCCAG	954
QY	796	GAAGCAAGGAAGATATAAGAGTCATTCAGAAAAATGCTCTCAACATCCACTTCATGTA	855
Db	955	GAAGCAAGAGGAGATATAAGAGTCTACTCAGAAAGCTCTCAGCACCCCTCCCATGTG	1014
QY	856	GAAGTATTACACTCAGAGATTATGGGTCTATCAGAAAAATTTGCTTTGGCTCTTGGTCTCGG	915
Db	1015	GAAGTCTGCACCTCCGTGGTTATGGCTCACCAGAAATTCGCTTCGCGCTTGGCTCTCGG	1074
QY	916	ATGAACAAAATTTAGACTATTCAAGTGACTTTAGGCAGATCTTTTGGCCAAAGCATGCC	975
Db	1075	ATGAACAAAATTTAGACTATTCAAGTGACTTTAGACAGATATTTTGGCAGGCTGCGCTC	1134
QY	976	AGAGAGAACTGACCTCGGAGATCCCTGTCTCATAAAGCAGGTTAATGCTTTGGGATGCA	1035
Db	1135	GTAGAAGAACTGGCTCTGAAAATCCCTGCTTTATGAAGAGACTAAATGCTTTGGGATGCA	1194
QY	1036	AAAGCTTTATAAGAGTGCCTTAAGATCCCTTCATGAAATTTGATCTTCAGCAGTTTTTTATG	1095

Db 2275 TTAAACAAGACCATATCCACAAGAACGAGATTGATTAAACGATATAATCATTTAATA 2334  
QY 2176 GAAGAAATGCTTCAGAGTCTCTCATATATTTGGGTGAGCCGTTATGTACCTGGAGTGGGA 2235  
Db 2335 GAAGAAATGCTTCAGAGTCTCTCATATATTTGGGTGAGAACGATATGTACCTGGAGTGGGA 2394  
QY 2236 AATGAGCAACAAGAGAGGTGCATATGAGAGAAATCATCTGCTTGGTTCATTTGAACCC 2295  
Db 2395 AATGTTACAGAGAGAGGTGTTATATGAGAGATTACTACCTTACTTTGCAATGAGCC 2454  
QY 2296 ATGCCACAAGTGCCTATGGCCAAAATTTACCTGGAATGAAATAATGAAACTGGCTTA 2355  
Db 2445 ATGCCACAGAGTGCCTATGGCCAGAAACCTACTGGAAGCAAAATATGAACTGGCTTA 2514  
QY 2356 GAGATGTCATTAACAAGTGGCCCATTTAAGAAACGAGTGTATCAGGCCATGAGATT 2415  
Db 2515 GAGATGTCATTAACAAGTGGCCCATTTAAGAAACGAGTGTGTGGGCAATGAGATT 2574  
QY 2416 TATGAACATAAAGATGATCACTGTAAGACTCAATATGTACTTTATCATTAACATCCAA 2475  
Db 2575 TATGATTTGAAGATGAATTCATGTAAGACTTCAATATGTACTTTTACCATTAATCTTAA 2634  
QY 2476 ACCAGCATAGCAAGGCTGAACATATGCAAGAAAGAGAAACAAGAAACAAGAT 2535  
Db 2635 ACACAGCATAGCAAGGCTGAACATATGCAAGAAAGAGAAACAAGAAACAAGAT 2694  
QY 2356 GAAGCATTTGCCGCCACCAACCTCTGTAATTTGCTGCTTTCAGCAAGTGAATAC 2595  
Db 2695 GAAGCATTTGCCGCCGCCACCTCTGTAAGTCTGCTGCTTTCAGCAAGTGAATAC 2754  
QY 2596 CTTCGACATGATGATATGATATGATATGATATGATATGATATGATATGATATGATATG 2655  
Db 2755 CTGCTGACATGATGATATGATATGATATGATATGATATGATATGATATGATATGATATG 2814  
QY 2656 ACAGATTCATCTGTTGACAGCAAGAGATGCTCCAAATGGCTTTTCATATTCGATG 2715  
Db 2815 ACAGATTCATCTGTTGACAGCAAGAGATGCTCCAAATGGCTTTTCATATTCGATG 2874  
QY 2716 GGTTTACTAGAGAGAGCAACAGCTTCAAAAAGCTCCTGAAGAAAGTAACTATTGAC 2775  
Db 2875 GGTTTACTAGAGAGAGCAACAGCTTCAAAAAGCTCCTGAAGAAAGTAACTATTGAC 2934  
QY 2776 TTTTATCTAAGGCTTCAAGATTTGGAGTTTCAAGCAT-----GAATATACAAAG 2826  
Db 2935 TTTTATCTAAGGCTTCAAGATTTGGAGTTTCAAGCATTTTCAAGATTTTCAAGATTTTCAAG 2994  
QY 2827 CTTTTGGAAAACTCAAGAAATTTCCAGTTAGAGAGCCAGCAAGACATGATTAACGTTG 2886  
Db 2995 CTCTTGGAAAACTCAAGAAATTTCCAGTTAGAGAGCCAGCAAGACATGATTAACGTTG 3054  
QY 2887 ATACTTCAGATGTTTGACACAGTGAAGCATTTAAGAGAAATCTTGTTAATTTAGCA 2946  
Db 3055 ATACTTCAGATGTTTGACACAGTGAAGCATTTAAGAGAAATCTTGTTAATTTAGCA 3114  
QY 2947 ACCCATCAGAGTGGGAATCTATTAGATGATGATGATTTACTCATGATTAAGAAAAAGCA 3006  
Db 3115 ACCCATCAGAGTGGGAATCTATTAGATGATGATGATTTACTCATGATTAAGAAAAAGCA 3174  
QY 3007 GAAGAAAAAGAAAAAGTGAAGCTGTAGGCTTACGCGCAAGAAATCATGCTGCATGATG 3066  
Db 3175 GAAGGGAAGAAAAAGTGAAGCTGTAGGCTTACGCGCAAGAAATCATGCTGCATGATG 3234  
QY 3067 TCTGCTTACAGAAAAATCTCATTTGAACCTCATTAACCTATGATGACAAATACATCAGAA 3126  
Db 3235 TCTGCTTACAGAAAAATCTCATTTGAACCTCATTAACCTATGATGATGATGATGATGATG 3294  
QY 3127 ATGCTGGGAAAGAGATTTCAATTTAGAGAGAGAGACACCCAGCATGACATCA 3186  
Db 3295 GTAAAGGGAAGAGAGATTTCAATTTAGAGAGAGAGACACCTGACATGATGATGATGATG 3354  
QY 3187 TCTGGAATTTGCTTTGGTCTTAACAGGGGCTCATTTGTTACTGAAAAGAGAGGCTGACG 3246  
Db 3355 TCTGGAATTTGCTTTGGGCTTAACAGGGGCTCATTTGTTACTGAAAAGAGAGGCTGACG 3414

QY 3247 TGCATCTTTGCCAAGAGAGAGAGAGTGAATAATAGAAAAATATGCGATGTATATCG 3306  
Db 3415 TGCATCTTTGCCAAGAGAGAGAGAGTGAATAATAGAAAAATATGCGATGTATATCG 3474  
QY 3307 GCTGTGTCAGAAATCTACTGCTTAAACCAGACAGAGGGGAAAAACCATATGAATCTCA 3366  
Db 3475 GCATGTGTGAGAAATCTACCGCCCTTAACCCAGCAGAGAGGAAAGCCGTGTGACACTTA 3534  
QY 3367 GGAGAAAGCCCTAGAACCCATTTTATGATTCAGAGCTTGGCATATGAACTTATACAGGA 3426  
Db 3535 GGAGAAAGCCCTAGAACCCATTTTATGATTCAGAGCTTGGCATATGAACTTATACAGGA 3594  
QY 3427 AGCTGTGCTATGTAATGCAAGCGAGTGTCTGGGAGAAATTTTGAAGCTGTACAGTG 3486  
Db 3595 AGCTGTGCTATGTAATGCAAGCGAGTGTGTGGGAGAAATTTTGAAGCTGTACAGTG 3654  
QY 3487 AGCTGTGAGAGGAGCATTCATGTTGACCTTTTACCTTTGGAAGTGGGAATATCTTTGC 3546  
Db 3655 AGCTGTGAGAGGAGCATTCATGTTGACCTTTTACCTTTGAGACCGGCGAGTACTATGC 3714  
QY 3547 CCTCTTTGCAAAATCTCTGTCAATACTGTATCCCATTTATCTTTGCAACCTCAAAAG 3606  
Db 3715 CCGCTGTGCAAGTCTCTGTCAACACTGTATATCCCATTCATCCCTTTGCAACCGAGAG 3774  
QY 3607 ATTAACAGTGAATGCAAGTCTCTGTCAACCTTTTACCTTTGCAACGCTGTATACG 3666  
Db 3775 ATCAACAGTGAATGCGAGGCTCTGTCAACCTTTTGAACCTTTGCGCCGCTGTATACG 3834  
QY 3667 ACTGTGTGGCGAATATGAGTTATATATATATATATATATATATATATATATATATATAT 3723  
Db 3835 ACTGTGTGGCGAATATGAGTTATATATATATATATATATATATATATATATATATATAT 3894  
QY 3724 ATTCTATTTTCTTTATCAAGAGATGGAGATTTCTATTTGAGATTTCATTCATCTG 3783  
Db 3895 GTTCGTGTCTTTATATCAAGAGATGGGAGATTTCACTTTTGAATTCATTCATCTG 3954  
QY 3784 AGTTTGGGCTTATGCTTGTGATTAATATTAATATGATCAAGAAAGTATTTCTC 3843  
Db 3955 AGTTTGGGCTTATGCTTGTGATTAATATTAATATGATCAAGAAAGTATTTCTC 4014  
QY 3844 TTTGCAACAACATTTAT 3903  
Db 4015 TTTGCAACAACATTTAT 4074  
QY 3904 GTCCCATGCTGACCTGAGACACCTGCTTCACTATCCAGCAATTTGAATTTATG 3963  
Db 4075 GTCCCATGATGACCTGAGACACCTGCTTCACTATCCAGCAATTTGAATTTATG 4134  
QY 3964 GGAGATGAAGAAAACTCTGTTTGAAGCACTTCAAAATAGGACATATGTCTGAAA 4023  
Db 4135 GGAGATGAAGAAAACTCTATTTGGAGCACTTCAAAATAGGACATATGTCTGAAA 4194  
QY 4024 GCATTTATGACAGTTTGACAGTTGACAGAGATTTACCTGTCTCAGGCTCATATCAGAAA 4083  
Db 4195 GCGCTATATGAGTTTGACAGTTGACAGAGGATTTACCTGCTCAGGCTCATATCAGAAA 4254  
QY 4084 CATCTGCTGCTTCTATATAGTTGCTTCTCCATCAATTAATAGAGATGATACACATGC 4143  
Db 4255 CATCTGCTGCTTCTATATAGTTGCTTCTCCATCACTGATGATAGAGATGATACACATGC 4314  
QY 4144 CTCTGCTTATATAGATGTTTCAATGTTTGGTGGGCTGTATAGCATTTCCATCTTTG 4203  
Db 4315 CTCTGCTTATATAGATGTTTCAATGTTTGGTGGGCTGTATAGCATTTCCATCTTTG 4374  
QY 4204 TATTTGAGATACCTGTTGATTTGACAGCTTTCTCACTAGTTCTTCTATACACCTT 4263  
Db 4375 TATTTGAGATACCTGTTGATTTGACAGCTTTCTCACTAGTTCTTCTATATACACCTT 4434  
QY 4264 TATCTCTTCATTTGATCACCATGCGCACATGCTTACAGATCACTTACAGTGAACA - 4321  
Db 4435 TATCTCTTCATTTGATCACCATGCGCACATGCTTACAGATCACTTACAGTGAACA 4494

QY 4322 -----CAGGCTACCCCTTGTCTCAGGTTCAAGAGACAGTGAAGAGGCTCATTC 4371  
Db 4495 GATCTGTCTCCAGGCGCGCTTGTCTGAGGTTGAAGAGATAGTGAAGAGGCTCGCTGT 4554  
QY 4372 GCATCTTCTTCTTTCGAGAAATTTCTCAATATACAAAGTGGCTCCATTTGGTGTGATATT 4431  
Db 4555 GCATCTGCTTCTTCTTGTGAAGTGTGCGACACAGACGCGCTCACTGGTGGTGTCT 4614  
QY 4432 CTTGGCTGTATTTTGGGTCTCAGTGAAGATGGCATCACCCCTTATCTTCCCTGTGCT 4491  
Db 4615 CCGGCTGTGTCTGTGGTCTCCCTTGAAGAGGCGCATCACCCCTTACCTCCCTGTGCT 4674  
QY 4492 GCATCTTCTTCTTCCACTATTTACTTGGGTAACTCCGCTGAGCACTGCATACCAATTCT 4551  
Db 4675 GCATCTTCTTCCACTATTTACTTGGGTAACTCCGCTGAGCACTGCATACCAATTCT 4734  
QY 4552 GCAGAGGAGTACAGTGCATCTGTAGTATCTATCTTACCTACAAATTTGTCTCTG 4611  
Db 4735 GCTGAAGGAGATTCAGTGCATCTGTAGTATCTATCTTACCCACAATTTGTCTCTG 4794  
QY 4612 CTCCTCCAGATATTTGGGATCTGTAAAGCCCTTCTCCAGAGCGGTGTCAGATCTCT 4671  
Db 4795 CTTTTCAGGAATATTTGGGATACCATTAAGCCCTTACTACAGAGGTGGTGTGAGATCTCT 4854  
QY 4672 GCCTTAACTTGTGTGAAGCAAAACACCGTGTGAGTACCTAGCACTGAGCAAAAT 4731  
Db 4855 GCCTTAACTTGTGTGAAGCAAAAGTGTGTGAGTACCTAGCACTGAGCAAAAT 4814  
QY 4732 AGTTTGATAGAGTCTCTGATGATAGTGTCTCTGATCAATCAAGCTTCTCAATTCAGG 4791  
Db 4915 AGTTTGATAGAGTCTCTGAGGACTACAGTGTCTTAAATCAGGCTTCTCAATTCAGG 4974  
QY 4792 TGCCCAAGTGTGATGATGAGCAAAATGATCTGTCTCTGATCTCTCTCTCTCTCTCT 4851  
Db 4975 TGCCCAAGTGTGATGATGAGCAAAATGATCTGTCTCTCTCTCTCTCTCTCTCTCTCT 5034  
QY 4852 ATACTATGTTCTCAGAACATTTCTGCCAGGAATTTGTGAACGGGAGAGGTGGAGCT 4911  
Db 5035 ATCTGTGTCTCAGAACATTTCTGCCAGGAATTTGTGAACGGGAGAGGTGGAGCG 5094  
QY 4912 TGCAATTTTCAACGACTTCACTGTGAGCGGAGTGTGATTTTCTTAAATACAGAGAA 4971  
Db 5095 TGCGTTTTCATGCGCTTCACTGTGAGTGTGAGTGTGATTTTCTTAAATACCGAGAA 5154  
QY 4972 TGCCGATGCTGTCTGTTGAAGTAAAGCCAGAGGCTGTGCTTATCCAGCTTCTTACTTG 5031  
Db 5155 TGCAAGGTGTCTGTGGAAGAAAGCCAGAGGCTGTGCTTACCCAGCCCTTACTTG 5214  
QY 5032 GATGAATATGGAACACAGACCTTGCCTGGAAGGGGCAACCCCTTCAATTTATCTCGT 5091  
Db 5215 GATGAATATGGAACACAGACCTTGCCTGGAAGGGGCAACCCCTTCAATTTATCTCGG 5274  
QY 5092 GAGCGGTATCGGAAGCTCCATTTGGTCTGGCAACACACTGCATTTATAGAAGAGATTGCT 5151  
Db 5275 GAGCGGTATCGGAAGCTCCATTTGGTCTGGCAACACACTGCATTTATAGAAGAGATTGCT 5334  
QY 5152 AGGAGCAAGAGACTAATCAGATGTTATTTGGATTTCAACTGGCAGTTACTGTGA 5205  
Db 5335 CGGAGCCAGGAGACTAATCAGATGTTATTTGGATTTCAACTGGCAGTTACTGTGA 5388

RESULT 4  
AAH14878  
ID AAH14878 standard; cDNA; 2550 BP.  
XX  
AC AAH14878;  
XX  
DT 26-JUN-2001 (first entry)  
XX  
XX Human cDNA sequence SEQ ID NO:12731.  
XX  
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

OS Homo sapiens.  
XX  
PN EF1074617-A2.  
XX  
PD 07-FEB-2001.  
XX  
XX 28-JUL-2000; 2000EP-0116126.  
XX  
PR 29-JUL-1999; 99JP-0248036.  
PR 27-AUG-1999; 99JP-0300253.  
PR 11-JAN-2000; 2000JP-0118776.  
PR 02-MAY-2000; 2000JP-0183767.  
PR 09-JUN-2000; 2000JP-0241899.  
XX  
PA (HELI-) HELIX RES INST.  
XX  
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX  
XX WPI; 2001-318749/34.  
XX  
PT Primer sets for synthesizing polynucleotides, particularly the 5602  
PT full-length cDNAs defined in the specification, and for the detection  
PT of the abnormality of the proteins encoded by the  
PT full-length cDNAs -  
XX  
PS Claim 8; SEQ ID 12731; 2537pp + CD ROM; English.  
XX  
CC The present invention describes primer sets for synthesizing 5602  
CC full-length cDNAs defined in the specification. Where a primer set  
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
CC to the complementary strand of a polynucleotide which comprises one of  
CC the 5602 nucleotide sequences defined in the specification, where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in  
CC the specification. The primer sets can be used in antisense therapy and  
CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
CC represent oligonucleotides, all of which are used in the exemplification  
CC of the present invention.  
XX  
SQ Sequence 2550 BP; 830 A; 454 C; 558 G; 708 T; 0 other;  
Query Match 46.8%; Score 2434.4; DB 22; Length 2550;  
Best Local Similarity 97.6%; Pred. No. 0;  
Matches 2488; Conservative 10; Mismatches 7; Indels 45; Gaps 1;  
QY 494 ATCATGAACCTGGAAGAGCAGGTACTATAAAGAGATTCACGCTGTCGTTGAATGAAG 553  
Db 1 ATCATGAACCTGGAAGAGCAGGTACTATAAAGAGATTCACGCTGTCGTTGAATGAAG 60  
QY 554 AGGTAATGTCACAGCCAGGAAATATTTCTTCTAGTGAATAATATGTCGTAGAAATGA 613  
Db 61 AGGTAATGTCACAGCCAGGAAATATTTCTTCTAGTGAATAATATGTCGTAGAAATGA 120  
QY 614 CTATATGGGAAGAGAAAGAACTGCTCTGAACTCCAGATAGGAGGAAATGAA 663  
Db 121 CTATATGGGAAGAGAAAGAACTGCTCTGAACTCCAGATAGGAGGAAATGAA 180  
QY 664 -----YYCVDHSHSYDHGTCATATACAGCC 688  
Db 181 GATACTATTGTCCTTTTCAATGATGAACACCATTCATATCACCACGTCATATACAGCC 240

OY	689	TACAAAGAGCTTGTAGCTGTAGCTCGAGAGGCCAGTTGCATACCACTGCCATTGACA	748
Db	241	TACAAAGAGCTTGTAGCTGTAGCTCGAGAGGCCAGTTGCATACCACTGCCATTGACA	300
OY	749	AAGAGGCGCTCGGGCTGTTTAAACGGGAGGTTATGCTTGTGCGAGAGCAAGAGAG	808
Db	301	AAGAGGCGCTCGGGCTGTTTAAACGGGAGGTTATGCTTGTGCGAGAGCAAGAGAG	360
OY	809	ATATTAAGAGCATTCAGAAAATGTCTCTCAACATCCACTCATGTAGAGTATTACACT	868
Db	361	ATATTAAGAGCATTCAGAAAATGTCTCTCAACATCCACTCATGTAGAGTATTACACT	420
OY	869	CAGAGATTATGGCTCATCAGAAAATTTGGTTGGCTGTGGTCTGGTGTGAACAAATTA	928
Db	421	CAGAGATTATGGCTCATCAGAAAATTTGGTTGGCTGTGGTCTGGTGTGAACAAATTA	480
OY	929	TGAGCTATTCAGTGCATTTAGGAGATCTTTTGCCAAAGCATGCTTAGAGAAAGCTG	988
Db	481	TGAGCTATTCAGTGCATTTAGGAGATCTTTTGCCAAAGCATGCTTAGAGAAAGCTG	540
OY	989	ACTGGGAGAAATCCCTGTCTCATTAAGCAGGTTAATGCTTTGGGATGCAAGCTTTATAAG	1048
Db	541	ACTGGGAGAAATCCCTGTCTCATTAAGCAGGTTAATGCTTTGGGATGCAAGCTTTATAAG	600
OY	1049	GTGCCCCGAAAGATCCTTCATGAATTTGATCTTACAGATTTTATGAGATGGAATATCA	1108
Db	601	GTGCCCCGAAAGATCCTTCATGAATTTGATCTTACAGATTTTATGAGATGGAATATCA	660
OY	1109	AAAAACTCTTGTGATGGAATTTGTGAAGTATTTAACAACCTCAGAAAANATATATCA	1168
Db	661	AAAAACTCTTGTGATGGAATTTGTGAAGTATTTAACAACCTCAGAAAANATATATCA	720
OY	1169	GTGATGATCATGACAGAAATATCTCTATATCTGACATTTGAGTTGAGTTACTGTTTC	1228
Db	721	GTGATGATCATGACAGAAATATCTCTATATCTGACATTTGAGTTGAGTTACTGTTTC	780
OY	1229	CTACTCGGCGAGCATCTTATTTGAAGAGAGATTTATCTGCTCATTTCTGAACACTC	1288
Db	781	CTACTCGGCGAGCATCTTATTTGAAGAGAGATTTATCTGCTCATTTCTGAACACTC	840
OY	1289	TGCTAGAAAGTTTATCTGAGTACTTGGACAGAGAAATTAATTAACCTCCAGGGTTATA	1348
Db	841	TGCTAGAAAGTTTATCTGAGTACTTGGACAGAGAAATTAATTAACCTCCAGGGTTATA	900
OY	1349	GCCAGGACAAATTTGGAGAGTATATGCAATATATGTACCTAAAGTATATCTGTATCA	1408
Db	901	GCCAGGACAAATTTGGAGAGTATATGCAATATATGTACCTAAAGTATATCTGTATCA	960
OY	1409	GCAAAACCCACAATATGACAGAGAAAGATTAGAATGCAAGTCTCTGGAAGTTTGTGACTT	1468
Db	961	GCAAAACCCACAATATGACAGAGAAAGATTAGAATGCAAGTCTCTGGAAGTTTGTGACTT	1020
OY	1469	TTTTGAAGATTTCTTACCTGTATGCAAGGAAATGGAAGAAATCCGAAGACAGTTGGGCAAC	1528
Db	1021	TTTTGAAGATTTCTTACCTGTATGCAAGGAAATGGAAGAAATCCGAAGACAGTTGGGCAAC	1080
OY	1529	ACATTGGAAGTGCATCTGATTTGGAGAGCTGCCATTGCTATTCACAGATGCATTTGAAGATA	1588
Db	1081	ACATTGGAAGTGCATCTGATTTGGAGAGCTGCCATTGCTATTCACAGATGCATTTGAAGATA	1140
OY	1589	TTTTTACTCATGTCCCAAGAGTGTGTGTCTGTATGAAGAAGCTTACTGTGGCTTATA	1648
Db	1141	TTTTTACTCATGTCCCAAGAGTGTGTGTCTGTATGAAGAAGCTTACTGTGGCTTATA	1200
OY	1649	AAGAAATGCAAAAGCTGTGATTTGAGGTGCATACCAGTTTATATTTCAAGTGAAGACAG	1708
Db	1201	AAGAAATGCAAAAGCTGTGATTTGAGGTGCATACCAGTTTATATTTCAAGTGAAGACAG	1260
OY	1709	TATGACATATGCTGTGGACATATGTTTGGAAAAGAGTCTACAGAGTATCTGAGATCTTG	1768
Db	1261	TATGACATATGCTGTGGACATATGTTTGGAAAAGAGTCTACAGAGTATCTGAGATCTTG	1320
OY	1769	TAAACATACATCTGCACCTCTCTAGGACCCCTTGCTGTGCTTCATGATAGCTTTAAGACAGC	1828

Db	1321	TAAGCATACATCTGGCACTCTCTAGACCCCTTGCGTCTTCATGTACGTTAAACATGCG	1380
QY	1829	TGGGTCGCTGTTTCAAGACTGCATGAAATTTGTCCTTTTGGAGACTTTTCAAGTAGAGTAC	1888
Db	1381	TGGGTCGCTGTTTCAAGACTGCATGAAATTTGTCCTTTTGGAGACTTTTCAAGTAGAGTAC	1440
QY	1889	TAGTGAATATCCCTTACTGTGTGTCGGTGTTGGTGGTGGCCAGGTGTCTGTAGATGTGCG	1948
Db	1441	TAGTGAATATCCCTTACTGTGTGTCGGTGTTGGTGGTGGCCAGGTGTCTGTAGATGTGCG	1500
QY	1949	GAAGAAATGAGACTGTCCTCTATTTAGCCAGGTGTTTATTACCAAGATGTTAAAGTGCAGAG	2008
Db	1501	GAAGAAATGAGACTGTCCTCTATTTAGCCAGGTGTTTATTACCAAGATGTTAAAGTGCAGAG	1560
QY	2009	AAGAAATGTATGATTAAGATATCATCATGCTTCAGATTTGGTGCAATCTTTAAATGATCCCA	2068
Db	1561	AAGAAATGTATGATTAAGATATCATCATGCTTCAGATTTGGTGCAATCTTTAAATGATCCCA	1620
QY	2069	ATAACTCTCTGTACTGTACTTCAGAGGTATGAACTTGGCAGGCTTTTAAACAAGACA	2128
Db	1621	ATAACTCTCTGTACTGTACTTCAGAGGTATGAACTTGGCAGGCTTTTAAACAAGACA	1680
QY	2129	TATCTACAAAAGACCAGATTTGATTAACAATATATATACACTATATAGAAATGCTTC	2188
Db	1681	TATCTACAAAAGACCAGATTTGATTAACAATATATATACACTATATAGAAATGCTTC	1744
QY	2189	AGTCTCATCTATATTTGTGGGTGAGCGTTATGTACCTGGAGTGGGAAATGTGACCAAG	2248
Db	1741	AGTCTCATCTATATTTGTGGGTGAGCGTTATGTACCTGGAGTGGGAAATGTGACCAAG	1800
QY	2249	AAGAGTACAAATGAGAAATCATTCACCTTGCTTTCATGTAACCCATGCGACACAGT	2308
Db	1801	AAGAGTACAAATGAGAAATCATTCACCTTGCTTTCATGTAACCCATGCGACACAGT	1860
QY	2309	CCATTGGCCAAAATTTACTCTGAGATGAAATTAATGAACGGCTTTAGAGATGTCATTA	2368
Db	1861	CCATTGGCCAAAATTTACTCTGAGATGAAATTAATGAACGGCTTTAGAGATGTCATTA	1922
QY	2369	ACAAAGTGGCCATTTTAAGAAACCCAGGTGATACAGGCCATGAGATTTATGAATTAAG	2428
Db	1921	ACAAAGTGGCCATTTTAAGAAACCCAGGTGATACAGGCCATGAGATTTATGAATTAAG	1980
QY	2429	ATGAATCACTGAAGACTTCAATATGTACTTTATCATTTACTCCAAAACCAAGCATAGCA	2488
Db	1981	ATGAATCACTGAAGACTTCAATATGTACTTTATCATTTACTCCAAAACCAAGCATAGCA	2040
QY	2489	AGCGTGAACATATGAGAAAGAAAGGAGAAACAAAGAACAAAGATAGCATGTGGCG	2548
Db	2041	AGCGTGAACATATGAGAAAGAAAGGAGAAACAAAGAACAAAGATAGCATGTGGCG	2100
QY	2549	CACACACACCTCTGTAATTTGCCCCGCTTTCAGCAAGGTATTAACCTTGCACCTGTG	2608
Db	2101	CACACACACCTCTGTAATTTGCCCCGCTTTCAGCAAGGTATTAACCTTGCACACCTGTG	2160
QY	2609	ATATCATGATGTACTTCTCAGAGCCGTATTTAGCGGGCAATAGACACAGATTTAACT	2668
Db	2161	ATATCATGATGTACTTCTCAGAGCCGTATTTAGCGGGCAATAGACACAGATTTAACT	2220
QY	2669	TGTGACCGGAGGATGCTCCAAATGGCTTTTCATTTCTGGCANTGGGTTTACTAGAGAG	2728
Db	2221	TGTGACCGGAGGATGCTCCAAATGGCTTTTCATTTCTGGCANTGGGTTTACTAGAGAG	2280
QY	2729	AGAACCAACAGCTTCAAAAAGCTCTGAAAGAAAGATTAACATTTTATCATTAAGG	2788
Db	2281	AGAACCAACAGCTTCAAAAAGCTCTGAAAGAAAGATTAACATTTTATCATTAAGG	2340
QY	2789	CTTCAAGATTGGGAAGTTCAGCCCATGATATACAAATCTTTTGGAAAACCTCAAGGAA	2848
Db	2341	CTTCAAGATTGGGAAGTTCAGCCCATGATATACAAATCTTTTGGAAAACCTCAAGGAA	2400
QY	2849	TTCCCAAGTTAGAAAGCCAGAAAGCATGATACGTGATACCTTGATGTGTTGACACAG	2908



Db 2401 TTCCCGAGTTAGAGCCAGAGGACATGATAACGTGGATCTTCAGATCTTTCAGACAG 2460

QY 2909 TGAAGCGATTAGAGAAAATCTTGTAAATTTAGCAACCATCAGGATCGGAATCTA 2968

Db 2461 TGAAGCGGTTAGAGAAAATCTTGTAAATTTAGCAACCATCAGGATCGGAATCTA 2520

QY 2969 TTAAGAAATGATGAGATTACTCATGATAAAG 2998

Db 2521 TTAAGAAATGATGAGATTACTCATGATAAAG 2550

RESULT 5

ABQ75898

ID ABQ75898 standard; cDNA; 1635 BP.

XX

AC ABQ75898;

XX

DT 17-OCT-2002 (first entry)

XX

XX Human ubiquitin relative protein 46.64 cDNA.

DE

XX Human; ubiquitin relative protein 46.64; tumour; inflammation;

KW immunological disease; haemopathy; human immunodeficiency virus; HIV;

KW gene; ss.

XX

OS Homo sapiens.

XX

FH Key Location/Qualifiers

FT 55..1329

FT CDS

FT /\*tag= a

FT /product= "ubiquitin relative protein 46.64"

XX

PN CN1339485-A.

XX

PD 13-MAR-2002.

XX

PF 23-AUG-2000; 2000CN-0119708.

XX

PR 23-AUG-2000; 2000CN-0119708.

XX

PA (BODE-) BODE GENE DEV CO LTD SHANGHAI.

XX

PI Mao Y, Xie Y;

XX

WPI; 2002-464069/50.

DR

DR P-PSDB; ABQ75898.

XX

PT New polypeptide-human ubiquitin relative protein 46.64 for

PT treating malignant tumours, inflammations, immunological diseases,

PT haemopathy and human immunodeficiency virus infection -

XX

PS Claim 6; Page 24-25 (disclosure); 33pp; Chinese.

XX

CC The present invention discloses a new kind of polypeptide, human

CC ubiquitin relative protein 46.64, polynucleotides for encoding this

CC polypeptide, and a DNA recombination process to produce the polypeptide.

CC The present invention also discloses the method of applying the

CC polypeptide in treating various diseases, such as malignant tumours,

CC inflammations, immunological diseases, haemopathy and human

CC immunodeficiency virus (HIV) infection. The current sequence represents

CC the human ubiquitin relative protein 46.64 cDNA.

XX

SQ Sequence 1635 BP; 529 A; 290 C; 344 G; 472 T; 0 other;

XX

Query Match 25.2%; Score 1311; DB 24; Length 1635;

Best Local Similarity 99.2%; Pred. No. 0;

Matches 1317; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1387 GACCTAAAGTATATCTCTGATCAGCAACCCACCAATATGACAGAAAGATTAAAGATGCAG 1446

Db 1 GACCTAAAGTATATCTCTGATCAGCAACCCACCAATATGACAGAAAGATTAAAGATGCAG 60

QY 1447 TTCCTTGAAGGTTTCGATCTTTTGAAGATTCTTACCTGTATGCGGGGAATGGAAGAA 1506

Db 61 TTCCTTGAAGGTTTCGATCTTTTGAAGATTCTTACCTGTATCGAGGAATGAAGAA 120

QY 1507 ATCCGAAGACAGGTTGGGCAACACATTTGAAGTGGATCCTGATTTGGAGGCTGCCATTCCT 1566

Db 121 ATCCGAAGACAGGTTGGGCAACACATTTGAAGTGGATCCTGATTTGGAGGCTGCCATTCCT 180

QY 1567 ATACAGATGCAATTGAAGAATATTTTACTCATGTTTCCAAAGAGTGTGCTGTGATGAA 1626

Db 181 ATACAGATGCAATTGAAGAATATTTTACTCATGTTTCCAAAGAGTGTGCTGTGATGAA 240

QY 1627 GAACCTTTACTTGTGGCTTATAAAGAATGTCACAAAGCTGTGATGAGTGCAGTACCAGT 1686

Db 241 GAACCTTTACTTGTGGCTTATAAAGAATGTCACAAAGCTGTGATGAGTGCAGTACCAGT 300

QY 1687 TTCATATCTAGTAGCAAGACAGTAGTACAATCGTGTGACATAGTTTGGAAACAAGTCC 1746

Db 301 TTCATATCTAGTAGCAAGACAGTAGTACAATCGTGTGACATAGTTTGGAAACAAGTCC 360

QY 1747 TACAGAGTATCTGAGGATCTTTGTAAGCATACATCGCCACTCTCTPAGGACCTTCTGCTGGT 1806

Db 361 TACAGAGTATCTGAGGATCTTTGTAAGCATACATCGCCACTCTCTPAGGACCTTCTGCTGGT 420

QY 1807 CTTTCATGTACGTTTAAAGCAGGCTGGGTGCTGTTTCAAGACTGCATGAATTTGTGCTTTT 1866

Db 421 CTTTCATGTACGTTTAAAGCAGGCTGGGTGCTGTTTCAAGACTGCATGAATTTGTGCTTTT 480

QY 1867 GAGGACTTTTCAAGTAGAGGTACTAGTGAATATCTTTACGTTGTCTGGTGTGCTGGTTCG 1926

Db 481 GAGGACTTTTCAAGTAGAGGTACTAGTGAATATCTTTACGTTGTCTGGTGTGCTGGTTCG 540

QY 1927 CAGGTTGTGCTGAGATGTGCGGAAGAAATGACTGTCTCTTATTAGGCAGGTGTTTAT 1986

Db 541 CAGGTTGTGCTGAGATGTGCGGAAGAAATGACTGTCTCTTATTAGGCAGGTGTTTAT 600

QY 1987 TACCAAGATGTTAAGTGCAGAGAGAAATGTATGATAAAGATATCATGCTTCAGATT 2046

Db 601 TACCAAGATGTTAAGTGCAGAGAGAAATGTATGATAAAGATATCATGCTTCAGATT 660

QY 2047 GGTGCATCTTTAATGGATCCCAATAGTTCTTGTACTGTGTTTACCTTTCAGAGGTATCAACT 2106

Db 661 GGTGCATCTTTAATGGATCCCAATAGTTCTTGTACTGTGTTTACCTTTCAGAGGTATCAACT 720

QY 2107/GCCGAGGCTTTTAAACAAGACCATATCTACAAAGACCAGGATTTCATTAAACAATATAT 2166

Db 721 GCCGAGGCTTTTAAACAAGACCATATCTACAAAGACCAGGATTTCATTAAACAATATAT 780

QY 2167 ACCTAATAGAGAAATGCTTTCAGGTCCTCATCTATATTGTTGGTGGAGCGTTATGACCT 2226

Db 781 ACCTAATAGAGAAATGCTTTCAGGTCCTCATCTATATTGTTGGTGGAGCGTTATGACCT 840

QY 2227 GGAGTGGAAATGTGACCAAGAGAGGTCAACATGAGAGAAATCATTCATCTGCTTTCG 2286

Db 841 GGAGTGGAAATGTGACCAAGAGAGGTCAACATGAGAGAAATCATTCATCTGCTTTCG 900

QY 2287 ATTCAACCCATGCCACACAGTGCCATTCGCCAAAATTTACCTGAGAATGAAATATGAA 2346

Db 901 ATTCAACCCATGCCACACAGTGCCATTCGCCAAAATTTACCTGAGAATGAAATATGAA 960

QY 2347 ACTGGCTTAGAGAATGTGATTAACAAGTGGCCACATTTAAGAAACCCAGGTGTATCAGCG 2406

Db 961 ACTGGCTTAGAGAATGTGATTAACAAGTGGCCACATTTAAGAAACCCAGGTGTATCAGCG 1020

QY 2407 CATGGAGTTTATGAACCTAAAAGATGAATCATCTGAAAGACTTCAATATGCTTTTATCAT 2466

Db 1021 CATGGAGTTTATGAACCTAAAAGATGAATCACTGAAAGACTTCGATATGCTTTTATCAT 1080

QY 2467 TACTCCAAAACCCAGCATAGCAAGGCTCAACATATGACAGAGAAAGAGGAGAAACAGAA 2526

Db 1081 TACTCCAAAACCCAGCATAGCAAGGCTCAACATATGACAGAGAAAGAGGAGAAACAGAA 1140

QY 2527 AACAAAGATGAAGCATTTGCCGCCACCACCTCTCTGAAATTCCTGCTTTCAGCAAA 2586



1893 GGAATATCCTTTACGTTGCTGCTGCTGGTTGGTTGCCAGGTTGTTGCTGAGATGTGCCGAAG 1952  
1177 AGAACCCCTCTTTAGATGCTCTGTTGTTGGTCCCAAGTACATGCCGGAATGTGGAGAAG 1236  
1953 AAATGGACTGCTCTTATTAGCCAGGTGTTTATTACCAAGATGTTAAAGTGCAGAGAAGA 2012  
1237 AAATGGTTCTCTCTAGTAACCCAGATTTATTACTACCAATAAATGTGAATGCAGACGTGA 1296  
2013 AATGTATGATAAGATATCATATGTTTCAGATTTGGTGCATCTTTAATGATGCCAATAA 2072  
1297 GATGTTTGAAGAAGATGTAGTAATGCTTCAGACAGGTGTCTCCATGATGGATCCAAATCA 1356  
2073 GTTCTGTGTTACTTCCAGAGGTATGAACTTGCCAGGCTTTTAAACAGACCATAATC 2132  
1357 TTTTCTGTGATGATCATGCTCAGCCGCTTTGAACTTTATCAGATTTTTCAGTACTCCAGACTA 1416  
2133 TACAAAAG-----ACCAGGATTTGATTAACAAATATATACACT 2171  
1417 TGGAAAAGATTTAGTTCTGAGATTACCCATAGGATGTTGTTTCAGCAGAACATATCTCT 1476  
2172 AATAGAAGAAATGCTTTCAGTCCCTCATCTATATTGTTGGGTGAGCGTTTATGTACCTGGAGT 2231  
1477 AATAGAAGAAATGCTATACCTCATATAATGCTTGTGTTGGAGAGAGATTTAGTCTCGGAGT 1536  
2232 GGGAAATGTACCAAGAAGAGGTCAATATGAGAGAAATCATTCACATGCTTTCATGTA 2291  
1537 TGGACAGGTAAATGCTACAGATGAAATCAAGCAGAGATTTATCCATCAGTTGAGTATCAA 1596  
2292 ACCATGCCACACAGTGCCTATGCCAAAATTTACCTGAGAATGAAAATATGAAACTGG 2351  
1597 GCCTATGGCTCATAGTGAATTTGGTAAAGTCTTTACCTGGAATGAGAACRAGGAGACTGG 1656  
2352 CTTAGAGAAATGCTATRAACAAAGTGGCCACATTTAAGAAACCGAGTGTATCAGCCCATGG 2411  
1657 CATGGAGAGTGAATCGAAGCAGTGGCCATTTCAAGAACTTGGATTACAGGACCGAGG 1716  
2412 AGTTTATGAACTAAAGATGAATCAGTGAAGACTTCAATATGATGACTTTTATCATTAATC 2471  
1717 CATGTATGAATGAAACAGAGATGTCGCAAGAGTTTCAACTGTGATTTCTATCACTTTTC 1776  
2472 CAAAACCCACATAGCAAGCTGAACATATGCAAGAAAGGAGAGAAACAAAGAAACAA 2531  
1777 AAGGCCAAGACGTCCAAAGCAGAAAGAACGCGACGGAATTTGAAAGACAAATATAGAGA 1836  
2532 AGATGAAGACATGCGCCACACACCTCCTGATTTCTGCCCTGTTTTCAGCAAAAGTAT 2591  
1837 AGATACAGACATCCCACTCCGGTGTGGCTCCATTTCTGCCCTCTGTTTGAAGCCCTGGT 1896  
2592 TAACCTTCTCACTGTGATATCATGATGTATCATTTCTCAGGACCGTATTTGAGCGGGCAAT 2651  
1897 TAACATTTTGCAGTTCAGATGTCATGTTGTGCATCATGGGAACAAATTTGCAATGGGCTGT 1956  
2652 AGACACAGATTTCAACTTTGGACCGAAGGATGCTCCAATGCTTTTTCATATCTCTGGC 2711  
1957 GGAACATAATTTGGATATGCCCTGGTCAGAGTCCATGCTGCAAGGGGTGTACATTTAATTTGG 2016  
2712 ATTGGGTTTACTAGAAGAGAACAAGCTTCAAAAAGCTCCTGAAG---AAGAAGTAAC 2768  
2017 CATGGCACTACAGAAGAAAACAAATTTAGAGAATGTACGGAAGCATGTAGTAAC 2076  
2769 ATTTGACTTTTATCATAGGCTTCAAGATTTGGGAAGTTCAGCCATGAAT-----AT 2819  
2077 ATTTACCTTTCACTCAGAAGATATCAAAACCTGGTGAAGCGCCAAAATTTCTCTAGCAT 2136  
2820 ACAAATGCTTTTGGAAAACATCAAGAAATTTCCCGAGTTAGAAGCCAGAGGACATGAT 2879  
2137 ACTAGCTATGCTGGAAACACTACAAAATGCTCCCTACCTAGAAGTCCACAAAGACATGAT 2196  
2880 AAGCTGATATCTTTCAGATGTTTGAACAGTGAAGCGATTAAGAGAAAATCTCTGTTTAAT 2939  
2197 TCGGTGATATTTGAAGACTTTTAAATGCTGTTTAAAGATGAGGAGA-----GTTACC 2250  
2940 TGTAGCAACCATCATCAGGATCGGAATCTATTAAGAATGATGAGATTACTCATGATAAGA 2999

2251 TACCAGTCCCCTGGCAGAGACAGAAAGGAACCAATATGGAAGAGAGTTCAAGGGACAAGA 2310  
3000 AAAAGCAGAAGCAAAAAGAAAGCTGTAAGCTGCTAGGCTTACATCCGCCAAGATCATGCG 3059  
2311 CAAAGCTGAGAGGAAGAAAGACAGAGATTGCCAGACTGCCAGAGAAAGATCATGCG 2370  
3060 TCAGATGTCCTGCTTACAGAAAAAATTCATTGAAACTCATAACTCATGTATGACATAC 3119  
2371 TCAGATGTCGTAATGTCAGCGCATTTTATTGATGAAAACAAAGAACTCTTTTCAGCAGAC 2430  
3120 ATCAGAAATGCTCGGAAAGAAGATTCATATGAGGAAGAGAGACCCCGCAGCAGCTCAG 3179  
2431 ATTAGAATGGAATGCTCAACTCTGCTGTTCTTGATCAT-----AGCCCTGTGGCTTC 2484  
3180 TGACTACTCTAGAATTTGCTTTGGGTCTTAAACGGGTCCATCTGTTACTTGAAAGGAGGT 3239  
2485 AGATATGACACTTACAGCACTGGGCCCGCCACAAACTCAGGTTCTCGAACAAGACAAT 2544  
3240 GCTGACGTGCATCTTTGCCAAGAAAGACAGAGGTGAAATAGAAAATAATATGCCATGCT 3299  
2545 CGTTACATGTTATTTGCTCAAGAGAGCAAGTTTAAAGTGGAAAGCAGGCAATGCT 2604  
3300 ATTATCGCCTGTGTCAGAAATCTACTGCCCTTAAACCAGCACAGGGGAAAACCCATAGA 3359  
2605 CTTGGCAGCATTTGTTTCAGAGATCAACTGTATTATCAAAAACACAGAAGTAAATTTATCA 2664  
3360 ACTCTCAGGAGAAGCCCTAGACCCACTTTTTCATGATCCAGACTTTGGCATATGGAACCTTA 3419  
2665 AGATCCAGAAAA---TATGATCCATTTATCATGCCCTGATCTGCTGTGGGAACACA 2721  
3420 TACAGGAAGCTGTGTCATGTAATCACGCAAGTGTGCTGGCAGAGATATTTTGAAGCTGT 3479  
2722 CACTAGTAGCTGTGGCACATTTATGCATGCCCATTTGTTGGCAAGGATTTTGTATCCCT 2781  
3480 ACAGCTGAGCTCTCAGCAGCGCATTCATGTTGACCTTTT-----TGACTTGA 3527  
2782 TCAAGCTAAAGAACAGCGAAGGCAACAGAGATTACGCTTTACATACAGAGTATGATGTAGA 2841  
3528 AAGTGGAGAATATCTTGGCCCTCTTGGCAATCTCTGTGCAATCTGTGCAATCTCCCATAT 3587  
2842 AAACGGAGAATTCCTTGGCCCTTGTGCAATGCTGTAGTAATCTGTTATTTCC----- 2895  
3588 TCCCTTTGCAACTCAAAAGATTAACAGTGAGAAATGCAGATGCTGTTGCTCAACTTTTGAC 3647  
2896 TCTGCTGCTGCTCCCAAGAAATATTTTAAACAACAGGTTTAAATTTTTCAGACCAACAAA 2955  
3648 CTTGCAAGGTTGATACAGACTGTTCTGCCAGAAATATCAGGTATATAATATAAGACATGC 3707  
2956 TCTGACTCAGTGGATTAGAACAAATATCTCAGCAAAATAAAGCAATTACAGTTTCTTTAGGAA 3015  
3708 TAAAGGAGAAAACCCCAATTCCTTATTTTCTTAAATCAAGGAATGGGAGATTTCTACTTTGA 3767  
3016 AGAAGAAAGTACTCTTAATATGCTCTACAAGAAATTCAGAAATGTGGATGAATTAACA 3075  
3768 GTTCCATTTCCATCTGAGTTTGGCGTTGAGTCTTCGATTAATATTTCAAAATAGCATCAA 3827  
3076 GCTCCCTGAAGGGTTTCAGGCCCTGATTTTCGCTTAAGATCCCTTATTTCTGAGAGCATAA 3135  
3828 GGAATGTTTATCTCTTTTGGCAACAATTTATAGAAATTTGATTTGAAAGTGGCCACTGA 3887  
3136 AGAAATGCTAACGACATTTGGAACCTGCTACCTACAGGTGGGACTAAAGGTTTATCCCCAA 3195  
3888 TGAAGGATCCCTCAGTCCCATCTGACCTGAGGACCTCGGCTTTTCTACTATCCAGGC 3947  
3196 TGAAGAGATCTCTGTTTCCCATTAATGTTGGGTAGCTCGGCGGTACACCATCCAAAG 3255  
3948 AATTGAAATCTATTTTGGAGATGAAGAAAACCTCTGTTTGGGAGCACTTCAAATAGGCA 4007  
3256 CATAGAAAGATTTTGGAGTGAAGATAAACCATTTGTTGTTGTTGTTTACCTTTGCAGACT 3315  
4008 GCATTAATGCTGAAGACATTAATGCAGTTTGCAGTTTGCAGAGATTTACCTGCTCTCA 4067

Db 3316 GGATGACTGCTTGTAGCTATTGACGAGATTTGCCGACGACCTGGACAGTGGCATCACT 3375  
OY 4068 GGTCCGTATGAGAAACATCTGGTTCGTCTTCTATCACTGTTCTTCCATACATAAATC 4127  
Db 3376 TTCACTGGTGAAGACATTTTGTAAACCTTTTGCATCAGCGGCGTAATGACAGCCA 3435  
OY 4128 AGAAGATACACCACCTCTTCTGTCTATAGATCTGTTTTCATAGTTTTGGTGGTGTGTT 4187  
Db 3436 TGAGAACTTCATGATATAGATATTGACATGTTTATGTTGTTGTTGGCTGTGCT 3495  
OY 4188 AGCATTCCTCATCTTGTATTTGGATGAGACCCCTGTATCTGACGCCCTTCTCAGTTAGTTC 4247  
Db 3496 TGCATTTCTCGTGGTGAAGT-----TCAGATTTTTCAGGGATGAGCCCT 3540  
OY 4248 TTCCATTAACACCTTTATCTCTCCATTTGATTCACATGGACACATCTTCAGATCT 4307  
Db 3541 TGGCACTGGAACCTTCATCTTTCATCTGTTTACTATAGGACACATATACAGATCTT 3600  
OY 4308 ACTTACAGTAGACACAGGCGCTTACC--CCTTGTCTAGGTTTCAAGAAAGACAGTGAAGGCC 4364  
Db 3601 ACTTACCTCATGTACAGAAAGAAATGCGATGATCAAGAAATCCCTTGTGAAGAAAGA 3660  
OY 4365 TCATTCGCCATCTTCTTCTTTCGAGAAATTTCTGATATATACAAAGTGGCTCCATTGGGTG 4424  
Db 3661 ATCAGCAGTTCTTGTGTATTAACACCTTCCACAGATTAACGGGAAAGTCCCTGAAGA 3720  
OY 4425 TGATATTCCTGGCTGCTATTTGTGGGTCTCACTGAAGATGGCATACCCTTATCTTGG 4484  
Db 3721 AATACCATCCGCTGGCATCTGTGAGAGAGTGTCAAGTGGATCATCTTCTCTGAA 3780  
OY 4485 CTGTGCTCATGTTTTCCTCACTATTTTACTTGGGGTAACTCCGCTGAGGAACATGCTATC 4544  
Db 3781 GGTCTCTGTTATTTTTCATCTTAAATGAGATCTCTTCCACCGACATTCGAAGT 3840  
OY 4545 CAATTCTGACAGAGAGAGTACAGTGCACCTGTGACTATCTATCTTACCTCAAAATTT 4604  
Db 3841 --TCTGGAACAAACCAATTTTGAACATTTATGTATCTTCTTCCCTCAAAACCACT 3897  
OY 4605 GTTCCGCTCTCCGGAATATTGGATCTGAAGGCCCTGTCCACAGGGCGTGTG 4664  
Db 3898 CATTTGCCCTTTTCAAGAAATAGTGAATTAATGAATTCAGTATGAAAGTTGGTGGCG 3957  
OY 4665 AGATCTGCTTACTTAACTGTTTGAAGCAAAAAACACCGTGTGAGTACCTTCAAGAA 4724  
Db 3958 TAACAGTGAAGTTAAAGATATCTGAAGGTGAAGAGATGCTATTAAGTATCCAGAGA 4017  
OY 4725 AAGAAATAGTTGATGAGAGCTTCTGATGACTATATAGCTCTCTCTGATCAAGCTTCTCA 4784  
Db 4018 ATCTAACAATTAATAAACCCTTCCAGAGGATTAACAGCAGCTTATTAACAAGCATCAA 4077  
OY 4785 TTTCAGGTGCCAGGTCTGCGAGATGATGAGCAAGCAATCCTGTCTCTGCTTTTTCG 4844  
Db 4078 TTTTCTGCTGCCGAATCAGGTGTGATTAAGAGCAAGGCCCAACTCTGTGCTTGTG 4137  
OY 4845 TGGGGCTTACTATGTTTTCAGAACATTTTGTGCGCAGAAATTTGTAACGGGGAAGAGT 4904  
Db 4138 CGGATCTCTGCTGTCTCCCAAGTACTGTCTGCCAGACTGAACTGGAAGGGAGAGATGT 4197  
OY 4905 TGGAGCTTGCATTTTTCAGCAGCACTTCACTGTGAGCCGAGTCTGCAATTTCTTAAAT 4964  
Db 4198 AGGAGCTGCAAGCTCACTACTCTGTGGCTGTGAGTGGCATCTCTGAGAGT 4257  
OY 4965 CAGAGATGCGAGAGTGTCTGTGTAAGTTAAAGCCAGAGAGGTGTGCTTCAAGTCC 5024  
Db 4258 ACGGGAATGTCAAGTGTCTATTTTAACTGTGCAAAACCAAAAGGCTGTTTATCTCTCC 4317  
OY 5025 TTACTTGTGATTAATGAGAAACAGACCTGTGCGTGAAGAGGGGGAACCCCTTATTT 5084  
Db 4318 TTACCTTGTGATTAATGAGGAGACCGACAGGAGCTCAAGCGGGAATTCCTTATCAATTT 4377  
OY 5085 ATCTGATGAGGCTATCGGAAGCTCATTTGCTGTGCAACACACTGCTATTAAGAGA 5144  
Db 4378 ATGCAAAAGCGATTCAGAGAAATTCAGAACTCTGTGCAACCAACAGTGTCAAGAGGA 4437

OY 5145 GATTGCTAGAGCCCAAGACATAATCAAGATTTATTGGATTCACACTGGCA 5195  
Db 4438 AATTGACATGACAGAGAACCAATCAGACACTGTTGGCATTTACTGGCA 4488  
RESULT 7  
AAK52693  
AAK52693 standard; cDNA; 6850 BP.  
XX  
AC AAK52693;  
XX  
DT 06-NOV-2001 (first entry)  
XX  
DE Human polynucleotide SEQ ID NO 2222.  
XX  
KW Human: cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
KW nervous system disorder; arthritis; inflammation; ss.  
XX  
OS Homo sapiens.  
XX  
PN W0200157190-A2.  
XX  
PD 09-AUG-2001.  
XX  
PE 05-FEB-2001; 2001WO-US04098.  
XX  
PR 03-FEB-2000; 2000US-0496914.  
XX  
PR 27-APR-2000; 2000US-0560875.  
XX  
PR 20-JUN-2000; 2000US-0598075.  
XX  
PR 19-JUL-2000; 2000US-0620325.  
XX  
PR 01-SEP-2000; 2000US-0654936.  
XX  
PR 15-SEP-2000; 2000US-0663551.  
XX  
PR 20-OCT-2000; 2000US-0693325.  
XX  
PR 30-NOV-2000; 2000US-0728422.  
XX  
PA (HYSEQ-) HYSEQ INC.  
XX  
PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;  
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;  
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;  
XX  
DR WPI: 2001-476283/51.  
XX  
PT P-PSDB; AAM79560.  
XX  
PS Claim 1; Page 4579-4580; 6221pp; English.  
XX  
CC The invention relates to polynucleotides (AAK51456-AAK53435) and the  
CC encoded polypeptides (AAM78323-AAW80302) that exhibit activity elating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation.  
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666  
CC (AAM80020) are omitted as the relevant pages from the sequence listing  
CC were missing at the time of publication.  
XX  
SQ Sequence 6850 BP; 2037 A; 1322 C; 1454 G; 2037 T; 0 other;  
Query Match 20.9%; Score 1088.4; DB 22; Length 6850;  
Best Local Similarity 55.9%; Pred. No. 5,3e-280;  
Matches 2409; Conservative 0; Mismatches 1806; Indels 97; Gaps 14;



Db 2371 CTCAGATGCTGAAATGACAGCGCATTTTATGATGAAACAAAGAACTTTTCAGCAGA 2430  
 QY 3119 CATCAGAAATGCTGGGAAAGAGATTCATTAATGAGGAGAGAGCACCACAGCATCA 3178  
 Db 2431 CATTAAGACGTGAGTGCCTCAACCTGCTGCTGTTGATCAT-----AGCCCTGTGCTT 2484  
 QY 3179 GTGACACTGCTAGAAATGCTTTGGGTCTTAAAGGGGTCCATCTGTTACTGAAAGAGAG 3238  
 Db 2485 CAGATATGACACTTACAGCACTGGGCCCGCACAACTCAGTTCTCGAACAAGACAAAT 2544  
 QY 3239 TCTGACGCTGATCCTTTGGCAGAGAAACAGAGGAGTGAATAATAGAAATATATGCTATG 3298  
 Db 2545 TCGTTACATGTAATGTTGTGCAAGAGAGACAGAAATGTAATGGAAGAGAGGCAATGG 2604  
 QY 3299 TATTATCGGCTGTGTCCAGAAATCTACTGCTTAAACCCAGCAGAGGGGAAAAACCATAG 3358  
 Db 2605 TCTTGGCAGCATTTGTTACAGAGATCAACGTATATCAAAAAACAGAAATTAATTTTC 2664  
 QY 3359 AACTTCAGAGAGAGCCCTAGACCCACTTTTCATGATCCAGACTTGGCATATGGAATTT 3418  
 Db 2665 AAGATCCAGAAAA---TATGATCCATTAATTCATGACCCCTGATCTGCTTGGGAACAC 2721  
 QY 3419 ATACAGAGAGCTGTGTCATGTAATGACGACAGTGTGCGAAGATATTTGAAGCTG 3478  
 Db 2722 ACATGATAGCTGTGGCACATTAATGATGATCCCATTTGTGGCAAAAGTATTTGATTCG 2781  
 QY 3479 TACAGCTGACTCTCAGCAGCGCATCATGTTGACCTTT-----TGACTTGG 3526  
 Db 2782 TTCAACCTTAAGAACAGCGAAGGCAAGAGATTAAGCTTACATACAGCATATGATGTAG 2841  
 QY 3527 AAAGTGAGAAATCTTTGGCCCTTTTGCATATCTGTGCATTAATCTGATGATCCCATTA 3586  
 Db 2842 AAAACGAGAAATCTTTGCCCTTTGTGAATGCTGAGTAATTAAGTATTAATCC----- 2896  
 QY 3587 TTCCCTTGCACCTCAAAAGATTAACAGTAGAATGATGATGCTGCTGCTCAACTTTGA 3646  
 Db 2897 -TCTGCTGCTCTCCAGAAATATTTTAAACACAGCTTAATTTTTCAGACCAACAA 2955  
 QY 3647 CCTGACGAGTGTATACAGACTGTTCTGCGCAGAAATATCAGGTTAATAATATAGACATG 3706  
 Db 2956 ATCTGACTAGTGTATAGAACATATCTCAGCAAAATAAAGCATTAACATTTCTTGA 3015  
 QY 3707 CTAAAGAGAAAAACCAATTCCTATTTCTTAAATCAAGAAATGGAGATTTCACTTTGG 3766  
 Db 3016 AAGAAAGAACTACTCTTAATATGCTCTACAAAGAAATTCAGAAATGTGATATTAAC 3075  
 QY 3767 AGTTCCATTCATCTGAGATTTGGGCTTGAATCTGATTAATATCAATATGATCA 3826  
 Db 3076 AGCTCCCTAAGGGTTACAGGCTGATTTTGTCTTAAGATCCCTTATTTCTGAGAGCAATA 3135  
 QY 3827 AGGAATATGTTATTTCTTTGGCCACAACATTTATGAAATTTGATTAAGATGGACCTG 3886  
 Db 3136 AAGAAATGCTAACGACATTTTGAATGTAATCTACACAGAGTGGGACTTAAAGTTCATCCCA 3195  
 QY 3887 ATGAAAGGATCCTCGAGTCCCATGCTGACAGACACCTGCGCTTCACTATCCAGG 3946  
 Db 3196 ATGAAAGGATCCTCGTGTCCCATATATGTTGGGTAGCTGCGCTACACCAATCCAAA 3255  
 QY 3947 CAATTGAAATCTATTGGGAGATGAAGAAACCTCTGTTGGAGCACTTCAAAATAGGC 4006  
 Db 3256 GCATTAAGAAATTTGAGTGAATGAATTAACCATTTGTTGGCTTTACCTTGAGAC 3315  
 QY 4007 AGCATATATGCTGGAAGCATTAATGAGTTTGGAGTTGACAGAGGATTAACCTGCTC 4066  
 Db 3316 TGGATGCTGTCTTAAAGTCAATGACGATTTGGCCGACACATGACAGTGGCANTAG 3375  
 QY 4067 AGCTCTGATACAGAAACATCTGTTGCTTATCAAGTGTGTTCTTCAACATTAAT 4126  
 Db 3376 TTTCATGTGTGACAGACATTTTGTAAACCTTTTGCATCAGTGTGCTAATACAGCC 3435  
 QY 4127 CAGAGATACACCATGCTCTGTCTATATAGATCTGTTCAATGTTTGGTGGTCTGTCT 4186  
 Db 3436 ATGAGAACTTCATGATATTAATGATTAACATGTTTCAATTAATGTTGGGCTTGGTGC 3495

QY 4187 TAGCATTCACATCCTTGTATTGGGATGACCTGTTGATCTGCACCTTCTTCAATTAGTT 4246  
 Db 3496 TTGCATTTCTGCGTTGGCAGTG-----TCAGATTTTTCAGGATACAGCC 3540  
 QY 4247 CTTCCTATTAACACCTTTATCTCTTCATTTGATGACCTTGGCAGACATGTTCAATGAC 4306  
 Db 3541 TTGGCAGTGGAGACCTTACATTTTCCATCTGTGTTACTATGTCACATCATATACATCT 3600  
 QY 4307 TACTTACATGACACAGCGCTTACC---CCTTGTGAGTTTCAAGAGACAGTGAAGAG 4363  
 Db 3601 TACTTACCTCATATGTAACAAGAAATGGATGATGATCAAGAAATCCCTTGTGAAGAG 3660  
 QY 4364 CTCAATCCGCAATCTTTCTTTGCGAAGAAATTTCTCAATATACAGTGGCTCCATTTGGT 4423  
 Db 3661 AATCAGCAGTCTTGTCTTTGTAATAAACACTTACACAGATATACGGAAGTGGCTGAAG 3720  
 QY 4424 GTGATATCCGCGGTATTTGTTGGGTCTCACTGAAAGAAATGACATCACCCCTTATCTTC 4483  
 Db 3721 AAATACCATCCGCGGTGATCTGTGAGAGTGTACAGACTGGAATCATATGCTTCTGA 3780  
 QY 4484 GCTGTGTCATTTGTTTTCACATTAATCTTGGGTAATCTCCGCTGAGAACTGATATA 4543  
 Db 3781 AGTGTCTGCTTATTTTTCATTAATGATGAGCTTCTTCCACCCGACATTCAG 3840  
 QY 4544 CCAATTCGCAAGAGAGATACAGTGCATCTGTAGTATCTATCTTACCTTACCAAT 4603  
 Db 3841 T---TCTGGAACAACCATTTTGAACATTTATGATGATATCTTCCCTTACCAACAC 3897  
 QY 4604 TGTTCCTCTCTTCCAGAAATTTTGGATACTGTAAAGCCCTTGTCCAGAGCGGTGTG 4663  
 Db 3898 TCATTTGCTTTTCAAAATATGATGATGATGATGATGATGATGATGATGATGATGATG 3957  
 QY 4664 CAGATCTGCTTACTTAACCTGTTGAAGCAAAAAACACCGGTGACAGTACCTAGAA 4723  
 Db 3958 GTAAAGCAAGTAAAGATATCTAGAAGTGAAGAGATGCTATTAACATATCAACATTC 4017  
 QY 4724 AAAGAAATATGTTTGAAGCTTCTGATGATGATGATGATGATGATGATGATGATGATG 4783  
 Db 4018 AATCTAACAAATTAATTAACCTTCCAGAGATTTACAGCGCTCATTAATCAAGCAATCA 4077  
 QY 4784 ATTTGAGTGGCCACGCTGTGCAATGATGATGATGATGATGATGATGATGATGATGATG 4843  
 Db 4078 ATTTCTGCTGCGGAAATCAGTGTGATTAAGACAGAGACCCCACTGTGTGTGTGT 4137  
 QY 4844 GTGGGCTATATGTTCTCTGAAACATTTGTCGACGAGAAATGTAACGCGGAGAGAG 4903  
 Db 4138 GCGGATCTCTGTGTCTCCAGACTTACTGCTGCCAGACTGAACTGGAAGGGGAGATG 4197  
 QY 4904 TTGAGCTTGCATTTTTCACGCACTTCACTGTGAGCCGAGCTGTGCATTTTCTTAAAA 4963  
 Db 4198 TAGAGCCTGCACACTCACACTCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4257  
 QY 4964 TCAGAAATGCGGAGTGTCTGTGTTGAAGTAAACCCAGAGCTGTGCTATTCACAGCTC 5023  
 Db 4258 TACGGGAATGTCAGTGTATTTTAAAGTGAAGGCAAAACCAAGGCTTTTAAATCTCTC 4317  
 QY 5024 CTTAATGATGAATATGAGAAACAGACCCGTGCGCTGGAAGAGGGGCAACCCCTTCAT 5083  
 Db 4318 CTTAATGATGAATATGAGAAACAGACCCGTGCGCTGGAAGAGGGGCAACCCCTTCAT 4377  
 QY 5084 TATCTGTCGAGCGGATGTCGAGCTTCATTTGCTGTGCAACACACTGCATTAATGAAG 5143  
 Db 4378 TATGCAAGAGAGATTAAGAAATGTCACAACTCTGTGCAACACAGAGTGCAGAGAG 4437  
 QY 5144 AGATTGTCAGAGACCAAGACTAATGATGATGATGATGATGATGATGATGATGATGATG 5195  
 Db 4438 AAATGTCAGTCACAGAGAAACCAATCAGACACTGTGTGATGATGATGATGATGATG 4489

RESULT 8  
 AAX03300  
 ID AAX03300 standard; DNA; 1001 bp.

XX AAX03300;  
AC 25-MAR-1999 (first entry)  
XX Partial cDNA encoding a human ubiquitin-protein ligase, Ubri.  
XX Ubiquitin-protein ligase; Ubri; human; ubiquitinylation; degradation;  
KW N-end rule pathway; stress-related muscle wasting; inhibitor; screen; ss.  
XX Homo sapiens.  
OS  
XX  
XX Key Location/Qualifiers  
FH 1..999  
FT /\*tag= a  
FT /product= Ubri  
FT /note= "partial CDC"  
XX  
XX US5861312-A.  
XX  
XX 19-JAN-1999.  
XX  
XX 02-DEC-1997; 97US-0982956.  
XX  
XX 02-DEC-1997; 97US-0982956.  
XX (CALY ) CALIFORNIA INST OF TECHNOLOGY.  
XX  
XX Kwon YT, Varshavsky A;  
XX WPI; 1999-130395/11.  
XX P-PSDB; AAW84353.  
XX  
XX Mouse and human Ubri cDNA - useful for producing recombinant Ubri  
XX polypeptides  
XX  
XX Claim 2; Columns 27-30; 18pp; English.  
XX  
XX The present sequence encodes a partial ubiquitin-protein ligase called  
XX Ubri. The Ubri enzymes are involved in protein ubiquitinylation and  
XX ultimate degradation through the N-end rule pathway and have been  
XX linked to stress-related muscle wasting. Recombinant Ubri polypeptides  
XX can be used to screen for inhibitors of muscle wasting when this is  
XX associated with the N-end rule pathway.  
XX  
XX Sequence 1001 BP; 363 A; 186 C; 205 G; 247 T; 0 other;  
XX  
XX Query Match 19.1%; Score 996.2; DB 20; Length 1001;  
XX Best Local Similarity 99.7%; Pred. No. 8.9e-256;  
XX Matches 998; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
Qy 2059 ATGGATCCCAATAAGTCTTCTTACTGTTACTTCAGAGGTATGAAGTTCGCCGAGGCTTTT 2118  
Db 1 ATGGATCCCAACAAGTCTTCTTACTGTTACTTCAGAGGTATGAAGTTCGCCGAGGCTTTT 60  
Qy 2119 AACAGACCATATACAAAAGACGAGGATTTGATTAACAATAATATACACTAATAGAA 2178  
Db 61 AACAGACCATATACAAAAGACGAGGATTTGATTAACAATAATATACACTAATAGAA 120  
Qy 2179 GAAATGCTTCAGGTCCTCATCTATATGTTGGGTGAGCGTTATGTACCTGGAGTGGGAAT 2238  
Db 121 GAAATGCTTCAGGTCCTCATCTATATGTTGGGTGAGCGTTATGTACCTGGAGTGGGAAT 180  
Qy 2239 GTGACCAAAAGAGGTCACAAATCAGAGAAATCATTGCTTGCATTGAACCCCATG 2298  
Db 181 GTGACCAAAAGAGGTCACAAATCAGAGAAATCATTGCTTGCATTGAACCCCATG 240  
Qy 2299 CCACACAGTGCATTTGCCAAAATTTTACCTGAGAAATGAAATGAACTGGCTTAGAG 2358  
Db 241 CCACACAGTGCATTTGCCAAAATTTTACCTGAGAAATGAAATGAACTGGCTTAGAG 300  
Qy 2359 AATGTCATAACAAAGTGGCCACATTTTAAGAAACAGGTGATCAGGCCATGGAGTTTAT 2418  
XX

Db 301 AATGTCATAACAAAGTGGCCACATTTAAGAAACACAGGTGTATCAGGCCATGGAGTTTAT 360  
Qy 2419 GAACATAAAGATGAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 2478  
Db 361 GAACATAAAGATGAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 420  
Qy 2479 CAGCATAGCAAGGCTGAACATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2538  
Db 421 CAGCATAGCAAGGCTGAACATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480  
Qy 2539 GCATTGCGGCCACACACCTCTGAAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2598  
Db 481 GCATTGCGGCCACACACCTCTGAAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540  
Qy 2599 CTCAACTGTGATATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2658  
Db 541 CTCAACTGTGATATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 600  
Qy 2659 GATTCTAACTTGTGGACCGAAGGATGCTCCAAATGGCTTTTCATATTTCTGGCATTTGGGT 2718  
Db 601 GATTCTAACTTGTGGACCGAAGGATGCTCCAAATGGCTTTTCATATTTCTGGCATTTGGGT 660  
Qy 2719 TTACTAG 2778  
Db 661 TTACTAG 720  
Qy 2779 TATCATAAGGCTTCAAGATTTGGAGAGTTCAGGCATGAATATACAAATGCTTTTGGAAAA 2838  
Db 721 TATCATAAGGCTTCAAGATTTGGAGAGTTCAGGCATGAATATACAAATGCTTTTGGAAAA 780  
Qy 2839 CTCAAAGGATTTCCCGAGTTAGAGCCGAGAGGACATGATACGTGGTACTTTCAGATG 2898  
Db 781 CTCAAAGGATTTCCCGAGTTAGAGCCGAGAGGACATGATACGTGGTACTTTCAGATG 840  
Qy 2899 TTTGACACAGTGAAGCGATTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2958  
Db 841 TTTGACACAGTGAAGCGATTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900  
Qy 2959 TCGGAATCTATTAAAGATGATGAGATTTACTCATGATATAAGAGAGAGAGAGAGAGAGAG 3018  
Db 901 TCGGAATCTATTAAAGATGATGAGATTTACTCATGATATAAGAGAGAGAGAGAGAGAGAG 960  
Qy 3019 AAAGCTGAAGCTGTAGGCTATGCGCAGAGAGATCATGGC 3059  
Db 961 AAAGCTGAAGCTGTAGGCTTTCATGCGCAGAGAGATCATGGC 1001  
RESULT 9  
AAC86934  
ID AAC86934 standard; cDNA; 1001 BP.  
XX  
AC AAC86934;  
XX  
DT 02-APR-2001 (first entry)  
XX  
DE Nucleotide sequence of a partial human Ubri protein.  
KW Ubri; E3-type protein; ubiquitin system; ubiquitin-protein ligase;  
KW N-end rule pathway; intracellular pathogen; Lysteria monocytogenes;  
KW Yersinia enterocolitica; muscle wasting; infection; ss.  
XX  
OS Homo sapiens.  
XX  
XX Key Location/Qualifiers  
FH 1..999  
FT /\*tag= a  
FT /product= "Ubri"  
FT /note= "partial sequence"  
XX  
XX US6159732-A.  
XX  
PD 12-DEC-2000.  
XX





CC	broad range of physiological activity, including immunisation against
CC	and inhibition of infections, allergies and cancer; regulation of tissue
CC	formation and repair; activin/inhibin activity; chemokine/cytokine
CC	activity; blood coagulation regulation; and receptor/ligand agonist
CC	or antagonist activity. The polypeptides can be used for prevention
CC	and treatment of disorders including infections by bacteria, yeasts and
CC	viruses (including HIV) and protozoa; metabolic and hormonal disorders;
CC	immune disorders (including severe combined immunodeficiency (SCID)
CC	and AIDS; thrombosis; cancer; and traumatic or surgical wounds.
XX	
SQ	Sequence 3327 BP; 1033 A; 666 C; 739 G; 889 T; 0 other;
	Query Match 15.6%; Score 812; DB 20; Length 3327;
	Best Local Similarity 55.3%; Pred. No. 3.3e-206;
	Matches 1855; Conservative 0; Mismatches 1410; Indels 87; Gaps 11;
QY	1892 TGGAAATACCTTTACGTTGTCTGCTGGTTGGTTGCCCAGGTTGTTGCTGAGATGTGCGAA 1951
DB	
DB	8 TAGAACAACCCCTCTTAGATGTCCTTTGCTGTGTCGCCAAGTACATGTCGGGAATCTGGAGAA 67
QY	1952 GAAATGCAGTCTCTCTATTATAGCCAGTGTTTATTACCAAGATGTTTAAGTCGAGAGAAG 2011
DB	
DB	68 GAAATGGTCTCTCTAGTAGTAACACAGATTATTATACCATTAATGTGAATGCAGACGGT 127
QY	2012 AAATGTATGATAAGATATATCATCATGCTTCAGATTTGGTGCATCTTTAATGGATCCCAATA 2071
DB	
DB	128 AGATGTTTGACAAGGATGTAGTAATGCTTCAGACAGGTGTCTCCATGATGATCCAAATC 187
QY	2072 AGTTCCTGTTACTGGTACTTTCAGAGGTATGAACCTTCCGAGGCTTTTAAACAAGCCATAT 2131
DB	
DB	188 ATTTCCTGATGATCATGTCTCAGCGCTTTGAACTTTATCAGATTTTTCAGTACTCCAGACT 247
QY	2132 CTACAAA-----GACCAGGATTTGATTAAACAATATAAATACAC 2170
DB	
DB	248 ATGAAAGAGATTTAGTCTGAGATTACCATTAAGGATGTTGTTACAGACAACTACTC 307
QY	2171 TAATAGAAGAAATGCTTCAGTGCTCTCATCTATATTGTGGGTGAGCGTTATGTACCTGGAG 2230
DB	
DB	308 TAATAGAAGAAATGCTATACCTATTATATGCTTGTGGAGAGAGATTTAGTCTCGGAG 367
QY	2231 TGGGAATGTGACCAAGAAGAGGTCACAATGAGAAATCATTTCACTTGCTTTGCAATTG 2290
DB	
DB	368 TTGCACAGGTAAATGCTTACAGATGAATCAAGCGAGAGATTATCCACTCAGTTGAGTATCA 427
QY	2291 AACCCATGCCACAGTGCCATTTGCCAAAATTTTCACTGAGATGAAATAATTAACAACTG 2350
DB	
DB	428 AGCCTATGGCTCATAGTGAATTGGTAAGTCTTTTACCTGAAGATGAGAACAGAGAGACT 487
QY	2351 GCCTTAGAGATGTCAATAACAAAGTGCCACATTTAAGAAACCAAGGTGTATCAGGCCATG 2410
DB	
DB	488 GCATGGAGGTGTAATCGAAGCAGTTTGCCCATTTCAAGAAACCTGGATTAAACAGGACGAG 547
QY	2411 GAGTTTATGAACATAAAGATGAATCACTGAAGACATTCATATATGTACTTTTATCATTTACT 2470
DB	
DB	548 GCATGTATGAACATAAACCAGATGTGCCAAAGAGTTCAACTGTGTTTCTATCACTTTT 607
QY	2471 CCAAAACCCAGCATAGCAAGCCTGAACATATGCAGAGAAAGAGGAAACACAGAAAACA 2530
DB	
DB	608 CAAGGCGAACAACAGTCCAGCGAGAGAACGCGCAACGGAATTTGAAAGAGCAAAATAGAG 667
QY	2531 AAGATGAAGCATTTGCCGCCACCACCCTCTCTGAATTTCTGCCCTGCTTTTCAGCAAAAGTGA 2590
DB	
DB	668 AAGATACAGACATCCCAACCTCCGGGTGTGCTCCATCTCTGCTTGTGGAAGCCTGG 727
QY	2591 TTAACCTTCTCAACTGTGATATCATATGATGTACATTTCTCAGGACCCGATATTTGAGCGGSCAA 2650
DB	
DB	728 TTAACATTTTGCAGTCAGATGTATGTTGTGTCATGTCATGGGAACAAATCTCTCAATGGGCTG 787
QY	2651 TAGACACAGATTTCTAACTTGTGGACCGAAGGATGTCTCAAATATGGGCTTTTTCATATTTCTGG 2710
DB	
DB	788 TGGAAACATAATGGATATGCCCTGCAGTGTCCAGTGTGCAAGGGGTGTACATTTAATTTG 847
QY	2711 CATTTGGGTTTACTAGAGAGAAGCAACAGCTTCTCAAAGAGCTCTCTGGAAG-----AAGAAGTAA 2767

Db	848	GCATGGCGACTACAAAGAGAAGAAAAACACATTTAGAGAAATGTCACGGGAAGACATGTAGTAA	907
Qy	2768	CATTGTGACTTTTATCATTAAGGCTTCAAGATTTGGAAGTTTGGGAAGTTTCAGCCCATGAAT	2818
Db	908	CATTGACCTTCACCTCAGACAGATATCAAAACCTGGTGAAGCGCCAAAATAATTCCTTAGCA	967
Qy	2819	TACAAATGCTTTTGGAAAAACTCAAAGGAATTTCCCAGTTTAGAAGGCCAAGAGCATGA	2878
Db	968	TACTAGCTATGCTGGAACACTACAAAATGCTCCCTACCTAGAAGTCCACAAAGACATGA	1027
Qy	2879	TAAAGTGGATCTCAGATGTTTGACACAGTGAAGCGCATTAAGAGAAAAATCTGTGTTTAA	2938
Db	1028	TTGCGTGGATATTGAAGACTTTTAAATGCTGTAAAAAGATCAGGAGA-----GTTTAC	1081
Qy	2939	TTGTAGCAACCACATCAGGATCGGAATCTATTAGAAATGATGAGATTACTCATGATAAAG	2998
Db	1082	CTACAGTCCCGTGGCAGACAGAGAAGAACCATTAATGGAAGAGATTCGAAGGACAAAG	1141
Qy	2999	AAAAAGCAGAACGAAAAAGAAAGCTGAAAGCTGTAGGCTACATCGCCAGAAAGTATGCG	3058
Db	1142	ACAAAGCTGAGAGGAAGAAAGCAGAGATTGCCAGACTGCCAGAGAAAAGATCATGG	1201
Qy	3059	CTCAGATGTGTGCTTACAGAAAACTTTCATGAAACTCATAACTCATGTATGACAATYA	3118
Db	1202	CTCAGATGTCTGAAATCGACGGCATTTTATGATGAAAAACAAGAACTCTTTCAGCAGA	1261
Qy	3119	CATCAGAAATGCCCTGGGNAAGAATTCATATGAGGAGGAAGAGACACCCAGCAGTCA	3178
Db	1262	CATTAGAACTGGATGGCTCAACCTGCTGTCTTGTATCA-----TAGCCCTGTGGCTT	1315
Qy	3179	GTGACTACTCTAGAAATTTGCTTTGGGTGCTTAAACGGGGTCCATCTGTTACTGAAAAAGAGG	3238
Db	1316	CAGATATGACACTTACAGCACTGGGCCCGCCACAACTCAGGTTCTCTGAACAAGACAAT	1375
Qy	3239	TGCTGACGTGCATCCTTTTGCCAAAGAAACAGAGAGGTGAAATAGAAAAATTAATGCCATGG	3298
Db	1376	TCGTTACATGTAATTGTGTCAAGAGGACAAAGAAGTTAAAGTGGAAAGCAGGCGCAATGG	1435
Qy	3299	TATTTATCGGCTGTGTCCAGAAATCTACTGCCCTTAACCCAGCACAGCGGGAACCCCATAG	3358
Db	1436	TCTTGGCAGCATTTGTTTCAGAGATCAACTGTATTATCAAAAACAGAAAGTAAATTATTC	1495
Qy	3359	AACCTCTCAGGAGAAGCCCTAGACCCACTTTTCATGGATCCAGACTTGGCATATGGAACCTT	3418
Db	1496	AAGATCCAGAAAAATAT---GATCCATATTATCATGCACCCCTGATCTCTTTGTGGAACAC	1552
Qy	3419	ATACAGGAAGCTGTGGTCATGTAATGACCCAGTGTGCTGGCAGAACTATTTTGAAGCTG	3478
Db	1553	ACACTAGTAGCTGTGGGCACATTTATGCATGCCCATTTGTTGSCAAAGTATTTTGTATCCG	1612
Qy	3479	TACAGCTGAGCTCTCAGCAGCGCATTTCAATGTTGACCTTTT-----TGACTTGG	3526
Db	1613	TTCAAGCTTAAAGAACAGCAAGGCAACAGAGATTACGCTTACATACAGAGCTATGATGTAG	1672
Qy	3527	AAAGTGGAGAATATCTTTGGCCCTCTTTGCAAACTCTGTGCAATCTGTATCCCAATTA	3586
Db	1673	AAAACGGAGAATTCCTTTGGCCCTTTGTGAATGCTTGTAGTAATACATGTTATTCCTCTGC	1732
Qy	3587	TTCCCTTTGCAACCTCAAAGATAAACAGTGAAGATGCAGATGCTTCTTGCTCAACTTTTGA	3646
Db	1733	TGCTTT-----CTCCAAGAAATATTTTAAACAACAGGTTAAATTTTTCAGACCAACCAA	1786
Qy	3647	CCCTGGCAGGTGGATACAGACTGTTCTGGCCAGAAATATCAGGTTTAAATATATAGACATG	3706
Db	1787	ATCTGACTCAGTGGATTAGAACAAATATCTCAGCAAAATAAAGCAATTAACAGTTTCTTAGA	1846
Qy	3707	CTAAGGAGAAACCCAAATTCCTATTTCTTTAATCAAGGAATGGGAGATTCCTACTTTGG	3766
Db	1847	AAGAAGAAAGTACTTCCCTAATAATGCCTCTACAAAAGAATTCAGAAAAATGTGGATGAATAC	1906
Qy	3767	AGTTCCCATTCCTCAGTTTTCGGCGTTGAGTCTCGCATTTAAATATTTCAAAATAGCATCA	3826

Db 1907 AGCTCCCTGAGAGGCTTCAGGCGCTGATTTGCTCTTAAGATCCCTTATTTGAGACGATTA 1966  
 QY 3827 AGGAAATGGTATTTCTCTTGGCCACAACATTTATAGATTTGATGAAAGTCCACCTG 3886  
 Db 1967 AAGAAATGCTTAAGAGATTTGGAATGCTACCTACCAAGGAGGAGTAAAGGTTTCAATCCCA 2026  
 QY 3887 ATGAAAGGATCTGAGAGTCCCATGCTACCTGAGACACCTGCTTTCATCTACAGG 3946  
 Db 2027 ATGAAAGAGATCTGCTGTTCCATATATGTTGGGGTAGCTGCGGCTACACATCCAAA 2086  
 QY 3947 CAATTGAAATCTATTGGAGAGAAAGAAACCTGTTTGGAGACCTTCAAAATAGGC 4006  
 Db 2087 GATAGAAAGAAATTTTGAAGTGAAGATTAACCATTTGTTGCTTTCCTTACCTTGACAC 2146  
 QY 4007 AGCATATATGCTTGAAGCATTAATGACATTTGACATTCACAGAGAGATTAACCTGCTC 4066  
 Db 2147 TGGATGACTGCTTGAAGTGAAGATTTGAGAGATTTGCGCCAGACACCTGACAGTGGCA 2206  
 QY 4067 AGGCTCTGATACAGAAACATCTGCTGCTTCTATACATTTGTTCTTCAACATTAAT 4126  
 Db 2207 TTTCAAGTGGTGCAGAGACATTTTGTAACTTTTTCATACATGCTGCTTAATGACAGCC 2266  
 QY 4127 CAGAGATACACCATGCTCTGCTATATGATCTGTTTATGTTGGGGTGGCTGTGT 4186  
 Db 2267 ATGAGAACTCTTCATGATTAATGATTTGACATGTTTCAATTTATTTGAGGCTTGTGTC 2326  
 QY 4187 TAGCATTCCTCATCTTGTATTTGGAGTACCTGTTTATGCTGACCTTCTTCACTAGTT 4246  
 Db 2327 TTGCATTTCTGCTGCTGAGT-----TAGGATTTTTCAGGATTAAGC 2371  
 QY 4247 CTTCCTATTAACCATCTTATCTCTTCCATTTGATACATGACATGCTTTCATGATAC 4306  
 Db 2372 TTGGAGACTGGAACCTTCATCTTCCATCTGTTTCACTGATGACACATCATCATGATCT 2431  
 QY 4307 TACTTACAGTACAC---AGGCTTACCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4363  
 Db 2432 TACTTACCTCATGTCAGAGAAATGCTGATGATTAACAAATCCCTTGTGAAG 2491  
 QY 4364 CTCAATCCGATCTCTTCTTCTTGCAGAAATTTCTCAATATACAAAGTGGCTCATGGGT 4423  
 Db 2492 AATCAGCAGTTCTTCTTGTATTAACACTTCACAGATTAACGGAAGTGGCTTGAAG 2551  
 QY 4424 GTGATATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4483  
 Db 2552 AAATACCAATCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2611  
 QY 4484 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4543  
 Db 2612 AGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2671  
 QY 4544 CCAATTCGAGAGAGAGATGACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4603  
 Db 2672 T---TCTGGAACAAGCCATTTTGAACATTTATGATGCTATCTTCTTCCCTAACAAACC 2728  
 QY 4604 TGTTCCTGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4663  
 Db 2729 TCATTTGCTCTTTCAGAGAAATGAGATTAATGATTAATGATTAATGATTAATGATTAATG 2788  
 QY 4664 CAGATCTGCTTACTAACTGTTGAGCAAAAACACCGTGGCTGAGTACCTGAGAA 4723  
 Db 2789 GTAAAGTGAAGTAAAGATATCTAGAAGTGAAGATGCTATTAATTAATTAATTAATTA 2848  
 QY 4724 AAAGAAATGATTTGATAGAGCTTCCGATGATTAATGCTGCTGCTGCTGCTGCTGCTGCTG 4783  
 Db 2849 AATCTAAACAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2908  
 QY 4784 ATTTGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4843  
 Db 2909 ATTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2968  
 QY 4844 GTGGGCTATATCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4903  
 Db 2969 GCGGATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3028

QY 4904 TTGAGCTTGCAATTTTTCACGCACTTACCTGAGAGCGGAGTCTGATTTTCTAATAA 4963  
 Db 3029 TAGAGCTGCAAGCTACACCTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3088  
 QY 4964 TCAGAGATCCGAGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5023  
 Db 3089 TACGGAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3148  
 QY 5024 CTTCCTGATGATTAAGAGAAACAGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5083  
 Db 3149 CTTCCTGATGATTAAGAGAAACAGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3208  
 QY 5084 TATCTGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5143  
 Db 3209 TATGCAAGAGCGATTCAGAGAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3268  
 QY 5144 AGATGCTGAGAGCCAGAGACTATACATGATTTATTTGATTCACCTGCA 5195  
 Db 3269 AAATGGACATGACAGAGAAAGCCATCAGACATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3320

RESULT 11  
 AAX35731  
 ID AAX35731 standard; cDNA; 3502 BP.  
 XX  
 AC AAX35731:  
 XX  
 DT 09-JUL-1999 (first entry)  
 XX  
 DE cDNA encoding a protein identified by the signal sequence trap method.  
 KW Signal sequence trap method; SSR method; immunisation; inhibition;  
 KW infection; allergy; cancer; regulation; tissue formation; tissue repair;  
 KW activin activity; inhibin activity; chemokine activity;  
 KW cytokine activity; blood coagulation regulation; agonist; antagonist;  
 KW metabolic disorder; hormonal disorder; immune disorder;  
 KW severe combined immunodeficiency; SCID; AIDS; thrombosis; cancer;  
 KW wound; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN W09918126-A1.  
 XX  
 PD 15-Apr-1999.  
 XX  
 PF 06-Oct-1998; 98WO-JP04514.  
 XX  
 PR 07-Oct-1997; 97JP-0274674.  
 XX  
 PA (ONOV ) ONO PHARM CO LTD.  
 XX  
 PI Fukushima D, Shibayama S, Tada H;  
 XX  
 DR WPI; 1999-277254/23.  
 DR P-PSDB; AAY02376.  
 XX  
 PT Polypeptides identified by the signal sequence trap method from a  
 PT human cDNA library  
 PS  
 XX  
 Claim 4; Page 170-178; 281pp; Japanese.  
 XX  
 AAX35694-X35747 represent cDNA sequences that encode novel polypeptides  
 (AAY02356-84) which are identified from a human placental cDNA library  
 CC by the signal sequence trap (SSR) method. The polypeptides have a  
 CC broad range of physiological activity, including immunisation against  
 CC and inhibition of infections, allergies and cancer; regulation of tissue  
 CC formation and repair; activin/inhibin activity; chemokine/cytokine  
 CC activity; blood coagulation regulation; and receptor/ligand agonist  
 CC or antagonist activity. The polypeptides can be used for prevention  
 CC and treatment of disorders including infections by bacteria, yeasts and  
 CC viruses (including HIV) and protozoa; metabolic and hormonal disorders;  
 CC immune disorders (including severe combined immunodeficiency (SCID))

CC and AIDS; thrombosis; cancer; and traumatic or surgical wounds.

Sequence 3502 BP; 1093 A; 690 C; 760 G; 959 T; 0 other;

Query Match 15.6%; Score 812; DB 20; Length 3502;  
Best Local Similarity 55.3%; Pred. No. 3.4e-206;  
Matches 1855; Conservative 0; Mismatches 1410; Indels 87; Gaps 11;

QY 1892 TGGAAATACCTTTACGTTGTCTGGTGTGGTTGCCCAGGTTGTTGCTGAGATGTGGCGAA 1951

63 TAGAACACCCCTCTTAGATGTCCTTGTCTGTGTGCCCAAGTACATGCCGGAATGTGGAGAA 122

QV 1952 GAAATGGACTGTCTCTTATTAGCCAGGTGTTTATTACCAAGATGTTAAGTGCAGAGAAG 2011

123 GAAATCGGGTTCTCTCTAGTAAACACATTATATCTCCATATATCTGAAATCCACACCTC 183

[illegible]

102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1041 1042 1043 1044 1045 1046 1047 1048 1049 1050 1051 1052 1053 1054 1055 1056 1057 1058 1059 1060 1061 1062 1063 1064 1065 1066 1067 1068 1069 1070 1071 1072 1073 1074 1075 1076 1077 1078 1079 1080 1081 1082 1083 1084 1085 1086 1087 1088 1089 1090 1091 1092 1093 1094 1095 1096 1097 1098 1099 11

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
84

[illegible][illegible][illegible]

DD  
 303 ATGGAAATAGATTAGTTCAGAGATTACCCATATAGGAGTGTGTTCAGCAGAGACAAATACATC 302

QY 2171 TAA TAG AAG AAA TGC CTC AGG GCC CAT CTA TAT TGT GGG TGA GCG TTA GTAC C TGG AG 2230

DB 363 TAAATAGGAAGAAATGCTATACCTCATTTATATAATGCTTGTGGAGAGAGAGATTTAGTCCCTGGAG 422

QY 2231 TGGGAAATGTGACCAAGAAGAGGTCAACAATGAGAGAAATCATTCACATTGCTTTGCATTG 2290

Db 423 TTGGACAGGTAATGCTACAGATGAAATCAAGCGAGAGATTATCCATCAGTTGAGTATCA 482

QY 2291 AACCCATGCCACACAGTGCCATTGCCAAAATTTACCTGAGAAATGAAACTG 2350

D**b** 483 AGCCTATGGCTCATAGTGAATTGGTAAAGTCTTTACCTGAAGATGAGAACAAAGGAGACTG 542

Qy 2351 GCTTAGAGAAATGTCATAAACAAGTGGCCACATTTAAGAAACCAGGTGTATCAGGCCATG 2410

Db 543 GCATGGAGAGTGTAATCGAAGCAGTTGCCCATTTCAAGAAACCTGGATTAAACAGGACGAG 602

QY 2411 GAGTTATGAACTAAAAGATGAATCACTGAAAGACTTCAATATGTACTTTTATCATTTACT 2470

Db 603 GCATGTATGAACTGAAACCAAGAAATGTGCCAAGAGTTCAACTTGTATTTCTATCACTTT 662

QY 2471 CCAAAACCCAGCATAGCAAGGCTGAAACATATGCAGAAGAAAAGGAGAAAACAAGAAAAACA 2530

Db 663 CAAGGGCAGAACAGTCCAAGGCAGAAAGCGCAACGGAAATTGAAAAGACAAAATAGAG 722

QY 2531 AAGATGAAGCATTGCCGCCACCACCCTCCTGAATCTGCCCTGCTTTCAGCAAAGTGA 2590

Db 723 AAGATACAGCACTCCCACCTCCGGTGTGCCCTCCATTCTGCCCTCTGTTTGAAGCCTGG 782

2591 TTAACCTTCTCAACTGTGATATCATGTATCATTTCTCAGGACCGTATTTTGAGCGGGCAA 2650

Db 783 TTAACATTTTGCAGTCAGATGTCATGTTGTGTCATCATGGGAACAATTCTGCAATGGGCTG 842

0V 2651 TACACACACACTTCCTAACCTCCTCCACCCACCCATCCCTCCCAATACTCTTCTCATTTCTCTG 2710

042 043 044 045 046 047 048 049 050 051 052 053 054 055 056 057 058 059 060 061 062 063 064 065 066 067 068 069 070 071 072 073 074 075 076 077 078 079 080 081 082 083 084 085 086 087 088 089 090 091 092 093 094 095 096 097 098 099 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1041 1042 1043 1044 1045 1046 1047 1048 1049 1050 1051 10

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
84

[illegible]

THE UNIVERSITY OF CHICAGO

[illegible]

DD 303 CATTACCTTCACTCAGATAGATATATCATTAATCCCTGGGATAGAGCGCCCAAAAAATTCCTCCCTATGCA TGTZ

QY 2819 TACAAATGCTTTTGGAAAAACTCAAAGGAATTCCTCCAGTTAGAAAGGCCAGAAAGGACATGA 2878

[illegible]

Db 2082 ATGAAGAGATCCGCTGTTCCCATATATGTGGGGTACGTGCGGTAACACATCCAA 2141  
 QY 3947 CAATTGAAATATTTGGAGATGAAGAAACCTCTGTTGGAGACTTCAAAATAGGC 4006  
 Db 2142 GCATGGAAGAAATTTGATGATGAAGATTAACCATTTGTTGGTCTTACCTTGGAC 2201  
 QY 4007 AGCATATATGTCGTAAGCATTAATGACAGTTGACAGTTGCACAGAGATTACCTGTCAC 4066  
 Db 2202 TGGATGACTGCTTATAGTATGACAGATTTGGCCGACACACTGACAGAGTGCATCAG 2261  
 QY 4067 AGTCTGATACAGAAACATCTGTTGCTTCTATACAGTTGTTCTTCAACATTAAT 4126  
 Db 2262 TTTGAGTGGTCAAGACATTTTGTAACTTTTGCATCAGTGGGCTTAATGACAGCC 2321  
 QY 4127 CAGAAGATACACACAGCTTCTGCTATAGATCTGTTCACTTTTGGTGGTCTGCT 4186  
 Db 2322 ATGAGAACTTCCATGATATATGATATGACATTTTCAATTAATGTTGGGCTTGGTCC 2381  
 QY 4187 TAGCATGCCATCCCTGATATGGAGTGAACCCGTTGATGACGCTTCTCAGTTAGTT 4246  
 Db 2382 TTGCATTTCTGCGTTGGACGTG-----TCAGATTTTTCAGGGATCAGCC 2426  
 QY 4247 CTTCCTATACACACCTTTATCTCTTCATTTGATCACCATGACACATGCTTACATAC 4306  
 Db 2427 TTGGCACTGAGACCTTCCATTTCCATCTGTTACTATGGACACATCATACAGATCT 2486  
 QY 4307 TACTTACAGTACAC---AGGCTACCCCTCTGCTGAGTTCAGAGACAGTGAAGAG 4363  
 Db 2487 TACTTACCTCATGTCAGAGAGAAATGGCATGATCAAGAAATCCCTCTTGAAGAG 2546  
 QY 4364 CTCATTCGGCATCTTCTTCTTGGAGAAATTTCTCAATATACAGTGGCTCCATGGGT 4423  
 Db 2547 AATCAGCAGTCTTCTGTTGATTAACACTTCACAGATATGAGGAAATGCTTGAAG 2606  
 QY 4424 GTGATATTCGCGGTGATTTGCTGCTCACTAAGATGACATCCACCTTATCTTC 4483  
 Db 2607 AAATACCATCCGCGTGGATCTGTGAGAGAGTGTAGAGCTGAATCATGCTTTCCTA 2666  
 QY 4484 GCTGCTGATGATGTTTTCATATTTACTTTGGGTAATCCGCGCTGAAGAACTGCATTA 4543  
 Db 2667 AGTGTCTGCTTATTTTTCATTAATTAAGTGGATCCCTCCACCCGATTCATCAG 2726  
 QY 4544 CCAATTCGCAAGAGAGATACAGTCACTGTAGCTATCTATCTTACTTACATCAATT 4603  
 Db 2727 T---TCCGGAACAAGCCATTTTGAACATTTATGAGTATCTTCCATCAACAACCC 2783  
 QY 4604 TGTTCCTCTCTCCAGGAATATGGAGTACGTGAAGGCCCTTCTCCAGAGCCGTG 4663  
 Db 2784 TCATTTGCTTTTCAAGAAATATGATATATGATTAATTAATTAATTAATTAATTAAT 2843  
 QY 4664 CAGATCCCTGCTTACTAATGTTTGAAGCAAAAACCCGTGTCAGAGTACCTTAGAA 4723  
 Db 2844 GTACAGTGAAGTATTAAGATATCTAGAAAGTGAAGAGATCTTATAGATATCCAGAG 2903  
 QY 4724 AAAGAAATAGTTGATAGAGTCTCTGATGACTATAGTCCCTCTGATCAAGCTTCTC 4783  
 Db 2904 AATCTAACAAATTAATAACCTTCCAGAGATTACAGAGCCTCATTAATCAAGATCCA 2963  
 QY 4784 ATTTCAGTGGCCCGCTGCTGAGATGATGAGCGAAAGACCTCTCTCTGCTTTTCT 4843  
 Db 2964 ATTTCTCTGCTCCGAATCAGTGGTGAATAGAGAGAGGCCCACTCTGCTTGTGT 3023  
 QY 4844 GTGGGCTATATGATGCTGAGACATTTGCTGACAGAAATTTGGAACGGGGAAGAG 4903  
 Db 3024 GCGATCTCTGCTGCTGCCAGAGATTACTGCTGCCAGACTAATCTGGAAGGGAGAGATG 3083  
 QY 4904 TTGGAAGTTCATTTTTCAGCACTTCACTGTGAGCGGAGTCTGCAATTTTCTTAA 4963  
 Db 3084 TAGAGGCTGACAGAGCTACACACTTCTCTGAGGCTGAGAGTGGGCACTTCTCGAGAG 3143  
 QY 4964 TCAGAGAAATGCGAGTGTCTGCTGTTAGAGTAAAGCCAGAGAGCTGTGCTATCCAGCTC 5023  
 Db 3144 TACGGAAATGTCAGTGTCTATTTTATGCTGCAAAACCAAGGCTGTTTATTTCTCTC 3203

QY 5024 CTACTTGATGATATGAGAAACACACCCCTGGCTGTAAGAGGGCAACCCCTTCATT 5083  
 Db 3204 CTTACCTTGTAGTACTATGGGAGACGACGAGGACTAGAGGGGAATCTTTTACATT 3263  
 QY 5084 TATCTCGTGAAGCGTATCGGAGAGCTCATTTGCTGGCAACACACTGATTAATGAAG 5143  
 Db 3264 TATGCAAAAGCGGATTTACAGAAATTCAGAAAGTCTGTGACACCAACAGAGTGCACAGAG 3323  
 QY 5144 AGATTGCTAGAGCGCAAGAGACTAATGATGTTATTTGATTCAACTGCA 5195  
 Db 3324 AATTTGACATGACACAGAGAGCCAAATCAGACACTGTGGATTGACTGGCA 3375  
 RESULT 12  
 AAA02411  
 ID AAA02411 standard; cDNA; 733 BP.  
 AC AAA02411;  
 XX  
 XX 19-MAY-2000 (first entry)  
 DT  
 XX  
 XX Human colon cancer cell line polynucleotide sequence SEQ ID NO:2402.  
 KW Human; colon cancer; tumour; diagnosis; gene expression product;  
 KW Probe; detection; cancerous state; metastasis; identification;  
 KW breast cancer; oestrogen receptor-positive breast cancer; therapy;  
 KW oestrogen receptor-negative breast cancer; lung cancer; ss.  
 KW  
 OS Homo sapiens.  
 XX  
 XX  
 PN WO958675-A2.  
 XX  
 PD 18-NOV-1999.  
 XX  
 PF 13-MAY-1999; 99MO-US10602.  
 XX  
 XX 14-MAY-1998; 98US-0085426.  
 PR 15-MAY-1998; 98US-0085537.  
 PR 15-MAY-1998; 98US-0085696.  
 PR 21-OCT-1998; 98US-0105234.  
 PR 27-OCT-1998; 98US-0105877.  
 XX  
 PA (CHIR ) CHIRON CORP.  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;  
 PI Reinhard C, Gliese K, Randazzo F, Kennedy GC, Pot D, Kassam A;  
 PI Lamson G, Drmanac R, Ctkvenjakov R, Dickson M, Drmanac S, Labat I;  
 PI Leshowitz D, Kita D, Garcia V, Jones LM, Stache-Crain B;  
 XX  
 DR WPI; 2000-126369/11.  
 XX  
 PT Polynucleotide library used to determine cancerous states of mammalian  
 PT cells -  
 PS  
 PS Claim 1: Page 958; 1097pp; English.  
 XX  
 CC AAA00010 to AAA02716 represent polynucleotides isolated from cDNA  
 CC libraries constructed from human colon cancer cell lines. The present  
 CC invention also describes a method of detecting differentially expressed  
 CC genes correlated with a cancerous state of a mammalian cell, comprising  
 CC detecting at least one\* differentially expressed gene product in a test  
 CC sample derived from a cell suspected of being cancerous, where detection  
 CC of the differentially expressed gene product is correlated with a  
 CC cancerous state of the cell from which the test sample was derived.  
 CC The polynucleotides sequences can be used in a method for detecting  
 CC differentially expressed genes correlated with a cancerous state of a  
 CC mammalian cell. The polynucleotides can also be used as probes for  
 CC detecting and mapping related genes. They can be used in diagnosis and  
 CC prognosis of diseases and disorders (e.g. identification of  
 CC pre-metastatic or metastatic cancerous states, stages of cancer, or  
 CC responsiveness of cancer to therapy). This is particularly for breast

CC	cancer, oestrogen receptor-positive breast cancer, oestrogen receptor-negative breast cancer, lung cancer, and colon cancer.
XX	
Sequence	733 BP; 218 A; 145 C; 152 G; 210 T; 8 other;
Query Match	12.58; Score 652.8; DB 21; Length 733;
Best Local Similarity	96.24; Pred. No. 5.9e-164;
Matches	688; Conservative 9; Mismatches 13; Indels 5; Gaps 3;
QY	664 YVCVNDHHSYDHGTCTATATACAGCCTACAAGAGCTCTTGCACGTGTGAGCTCGCAGAGGCC 723
DB	19 CCAGGAGCACGTCATATACAGCCTACAAGAGCTCTTGCACGTGTGAGCTCGCAGAGGCC 78
QY	724 CAGTTCGATACCACTGCCATTGACAAAGAGGGTCGTGGGCTGTTAAAGCGGAGCTTAT 783
DB	79 CAGTTGCTNCTCCACTGCCATTGACAAAGAGGGTCGTGGGCTGTTAAAGCGGAGCTTAT 138
QY	784 GCTGCTGCCAGGAAGCAAGGAAGATATATAAGAGATCATTAAGAGATGCTCTCAACAT 843
DB	139 GCTGCTGCCAGGAAGCAAGGAAGATATATAAGAGATCATTAAGAGATGCTCTCAACAT 198
QY	844 CCACCTTCATGTGAAGTATTACACTCAGAGATATTGGCTCATCAGAAATTTGCTTTCGCT 903
DB	199 CCACCTTCATGTGAAGTATTACACTCAGAGATATTGGCTCATCAGAAATTTGCTTTCGCT 258
QY	904 CTTTGGTTCCTGSGATGAACAAAAATTTATGAGCTATTCAAGTGACTTTAGGCAGATCTTTTGC 963
DB	259 CTTGG -TCCTGGATGAACAAAAATTTATGAGCTATTCAAGTGACTTTAGGCAGATCTTTTGC 317
QY	964 CAAGCATGCTTTAGAGAGAACCCTGACTCGGAGAACCCCTGCTCATCAAGCAGGTTAATG 1023
DB	318 CAAGCATGCTTTAGAGAGAACCCTGACTCGGAGAACCCCTGCTCATCAAGCAGGTTAATG 377
QY	1024 CTTTGGGATGCAAGCTTTATAAGGTGCCCGTAAGATCCTTCATGAATTTGATCTTCAGC 1083
DB	378 CTTTGGGATGCAAGCTTTATAAGGTGCCCGTAAGATCCTTCATGAATTTGATCTTCAGC 437
QY	1084 AGTTTTTTTATGGAGATGGAATACAAAAACTCTTTGCTATGGAATTTGTGAAGTATTAT 1143
DB	438 AGTTTTTTTATGGAGATGGAATACAAAAACTCTTTGCTATGGAATTTGTGAAGTATTAT 497
QY	1144 AACCACTGCAGAAAGAAATATACAGTGATGATCATGACAGAGTAGTCTCTATAACWGCA 1203
DB	498 AACCACTGCAGAAAGAAATATACAGTGATGATCATGACAGAGTAGTCTCTATAACWGCA 557
QY	1204 CTTTCAGTTCAGATGTTTACTGTTCTCTACCTCGCTCGACATCTTATTGAAGACAGCAAT 1263
DB	558 CTTTCAGTTCAGATGTTTACTGTTCTCTACCTCGCTCGACATCTTATTGAAGACAGCAAT 617
QY	1264 GTTATCTCTCTATTACTGAAACTCTGCTAGAAGTTTTTACCTGAGTACTTTGGACAGG -A 1321
DB	618 GTTATCTCTCTATTACTGAAACTCTGCTAGAAGTTTTTACCTGAGTACTTTGGACNNGAA 677
QY	1322 ACAATAAATTCACCTCCAGGGTTATAGCC -AGGACAAATTTGGGAAGAGTATAT 1374
DB	678 CCAATAAATTCACCTCCAGGGTTATAGCCAGGACAAAATTTGGGAAGAGTATAT 732
RESULT	13
AA02327	
ID	AAA02327 standard; cdna; 756 BP.
XX	
AC	AAA02327;
XX	
DT	19-MAY-2000 (first entry)
XX	
DE	Human colon cancer cell line polynucleotide sequence SEQ ID NO:2318.
XX	
KW	Human; colon cancer; tumour; diagnosis; gene expression product;
KW	probe; detection; cancerous state; metastasis; identification;
KW	breast cancer; oestrogen receptor-positive breast cancer; therapy;
KW	oestrogen receptor-negative breast cancer; lung cancer; ss.
XX	

QY 961 TGGCAAGCATGCTTGAAGAAACCGTACTGGAGATCCCTGTCATTAAGCAGGTGA 1020  
 DB 350 TGGCAAGCATGCTTGAAGAAACCGTACTGGAGATCCCTGTCATTAAGCAGGTGA 409  
 QY 1021 ATGCTTTGGGATCAAAAGCTTTTAAAGGTGCCCGTAAAGTCCCTTCAATGATGATCTTC 1080  
 DB 410 ATGCTTTGGGATCAAAAGCTTTTAAAGGTGCCCGTAAAGTCCCTTCAATGATGATCTTC 469  
 QY 1081 AGCAGTTTTTTTATGAGATGAGTAACAAAACTCTTCTGATGAAATTTGGAGATAT 1140  
 DB 470 AGCAGTTTTTTTATGAGATGAGTAACAAAACTCTTCTGATGAAATTTGGAGATAT 529  
 QY 1141 TATAACAACCTGAGAGAAATATATCAGTATGATGATGATGATGATGATGATGATGAT 1200  
 DB 530 TATAACAACCTGAGAGAAATATATCAGTATGATGATGATGATGATGATGATGATGAT 589  
 QY 1201 GCACCTTCAGTTCAGATGTTTACTGCTCTACTCTGCTGCTGCTGCTGCTGCTGCTGCT 1260  
 DB 590 GCAC-CTTCAGTTCAGATGTTTACTGCTCTACTCTGCTGCTGCTGCTGCTGCTGCTGCT 647  
 QY 1261 AATGTTATCTCTGCTCTACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1320  
 DB 648 AATGTTATCTCTGCTCTACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 707  
 QY 1321 AACAAATTAATCAACTCCAGGCTTATAGCCAGAGACAATTTGGAGAGAT 1370  
 DB 708 AACAAATTAATCAACTCCAGGCTTATAGCCAGAGACAATTTGGAGAGAT 756

RESULT 14  
 AAH07621  
 ID AAH07621 standard; cDNA; 712 BP.  
 XX  
 AC AAH07621;  
 XX  
 DT 26-JUN-2001 (first entry)  
 XX  
 DE Human cDNA clone (5'-primer) SEQ ID NO:4456.  
 XX  
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
 XX  
 OS Homo sapiens.  
 PN EP1074617-A2.  
 PD 07-FEB-2001.  
 XX  
 PF 28-JUL-2000; 2000EP-0116126.  
 XX  
 PR 29-JUL-1999; 99JP-0248036.  
 PR 27-AUG-1999; 99JP-0300253.  
 PR 11-JAN-2000; 2000JP-0118776.  
 PR 02-MAY-2000; 2000JP-0183767.  
 PR 09-JUN-2000; 2000JP-0241899.  
 XX  
 PA (HELI-) HELIX RES INST.  
 PI Ota T, Isegai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 XX  
 DR WPI; 2001-318749/34.  
 XX  
 PT Primer sets for synthesizing polynucleotides, particularly the 5602  
 PT full-length cDNAs defined in the specification, and for the detection  
 PT and/or diagnosis of the abnormality of the proteins encoded by the  
 PT full-length cDNAs -  
 XX  
 PS Claim 1; SEQ ID 4456; 2537bp + CD ROM; English.  
 CC  
 CC The present invention describes primer sets for synthesizing 5602  
 CC full-length cDNAs defined in the specification. Where a primer set  
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
 CC to the complementary strand of a polynucleotide which comprises one of

CC the 5602 nucleotide sequences defined in the specification, where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to the  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in  
 CC the specification. The primer sets can be used in antisense therapy and  
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
 CC represent oligonucleotides, all of which are used in the exemplification  
 CC of the present invention.  
 XX  
 SQ Sequence 712 BP; 229 A; 132 C; 156 G; 190 T; 5 other;

Query Match 10.3%; Score 534.6; DB 22; Length 712;  
 Best Local Similarity 89.1%; Pred. No. 2.5e-132;  
 Matches 627; Conservative 10; Mismatches 18; Indels 49; Gaps 4;

QY 500 AACCTGGAAGCAGAGTACTATATAAAGAAATTCACGCTGCTGCTGAATGAAGAGTAA 559  
 DB 8 ACTGGGAAAGCAGAGTACTATATAAAGAAATTCACGCTGCTGCTGAATGAAGAGTAA 67  
 QY 560 TTGTCCAAAGCAGGAAATATTTCTCTAGTATATAATATGTCGTAGAAATGACTATAT 619  
 DB 68 TTGTCCAAAGCAGGAAATATTTCTCTAGTATATAATATGTCGTAGAAATGACTATAT 127  
 QY 620 GGAAGAAGGAAAGAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 663  
 DB 128 GGAAGAAGGAAAGAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 187  
 QY 664 -----YYCVNDHSHSYDHSTCATATACAGCTTCAAA 694  
 DB 188 ATGTGTCCTTTTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 247  
 QY 695 GAGCTCTTGACTGTGAGCTGCGAGAGCCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 754  
 DB 248 GAGCTCTTGACTGTGAGCTGCGAGAGCCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 307  
 QY 755 GTGCTGGGCTTAAAGCGGAGGCTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 814  
 DB 308 GTGCTGGGCTTAAAGCGGAGGCTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 367  
 QY 815 AGAGTCATTCAGAAATGCTCTCAACATCCACTTCATGATGATGATGATGATGATGATGATGAT 874  
 DB 368 AGAGTCATTCAGAAATGCTCTCAACATCCACTTCATGATGATGATGATGATGATGATGATGAT 427  
 QY 875 TTATGCTCATCAGAAATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 934  
 DB 428 TTATGCTCATCAGAAATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 487  
 QY 935 ATTCAATGACTTTAGAGAGATCTTTTCCCAAGCATGCTTGAAGAAGAACTGACTCGG 994  
 DB 488 ATTCAATGACTTTAGAGAGATCTTTTCCCAAGCATGCTTGAAGAAGAACTGACTCGG 547  
 QY 995 AGAATCCCTGCTCATAGAGAGGTTAATGCTTTGGAGATGCAAAAGCTTATAAAGG-CC 1053  
 DB 548 AGAATCCCTGCTCATAGAGAGGTTAATGCTTTGGAGATGCAAAAGCTTATAAAGGTTGCC 607  
 QY 1054 CGTAAAGTCCCTCATAGATGATCTTCAGAGTTTATAGG-AGATGGAATACAAAA 1112  
 DB 608 CGTAAAGTCCCTCATAGATGATCTTCAGAGTTTATAGG-AGATGGAATACAAAA 667  
 QY 1113 ACTCTTGCATGGAATTT--GTGAAGTATTATAAACAAGTCA 1154  
 DB 668 ACTCTTGCATGGAATTTGTGAAGTATTATAAACAAGTCA 711



RESULT 15  
AAS70068  
ID AAS70068 standard; cDNA; 777 BP.  
XX AC AAS70068;  
XX AC AAS70068;  
DT 13-FEB-2002 (first entry)  
XX DE DNA encoding novel human diagnostic protein #5872.  
XX DE Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
KW Homo sapiens.  
XX OS  
XX WO200175067-A2.  
XX PN  
XX PD 11-OCT-2001.  
XX PF 30-MAR-2001; 2001WO-US08631.  
XX PR 31-MAR-2000; 2000US-0540217.  
XX PR 23-AUG-2000; 2000US-0649167.  
XX XX (HYSE-) HYSEQ INC.  
XX XX Drmanac RT, Liu C, Tang YT;  
XX PI  
XX DR WPI; 2001-639362/73.  
XX DR P-PSDB; ABG05881.  
XX XX  
XX PT New isolated polynucleotide and encoded polypeptides, useful in  
XX PT diagnostics, forensics, gene mapping, identification of mutations  
XX PT responsible for genetic disorders or other traits and to assess  
XX PT biodiversity  
XX XX  
XX FS Claim 1; SEQ ID No 5872; 103pp; English.  
XX XX  
XX CC The invention relates to isolated polynucleotide (I) and  
XX CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
XX CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
XX CC and gene mapping, and in recombinant production of (II). The  
XX CC polynucleotides are also used in diagnostics as expressed sequence tags  
XX CC for identifying expressed genes. (I) is useful in gene therapy techniques  
XX CC to restore normal activity of (II) or to treat disease states involving  
XX CC (II). (II) is useful for generating antibodies against it, detecting or  
XX CC quantitating a polypeptide in tissue, as molecular weight markers and as  
XX CC a food supplement. (II) and its binding partners are useful in medical  
XX CC imaging of sites expressing (II). (I) and (II) are useful for treating  
XX CC disorders involving aberrant protein expression or biological activity.  
XX CC The polypeptide and polynucleotide sequences have applications in  
XX CC diagnostics, forensics, gene mapping, identification of mutations  
XX CC responsible for genetic disorders or other traits to assess biodiversity  
XX CC and to produce other types of data and products dependent on DNA and  
XX CC amino acid sequences. AAS64197-AAS94564 represent novel human  
XX CC diagnostic coding sequences of the invention.  
XX CC Note: The sequence data for this patent did not appear in the printed  
XX CC specification, but was obtained in electronic format directly from WIPO  
XX CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX SQ Sequence 777 BP; 194 A; 204 C; 199 G; 180 T; 0 other;

Query Match 8.7%; Score 450.8; DB 23; Length 777;  
Best Local Similarity 99.6%; Fred. No. 6.9e-110;  
Matches 452; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 3165 CACCCACGACGACTACTCTAGAAATGCTTGGGTCTCTAAACGGGGTCCATCTGT 3224  
|||||  
Db 324 CACCCACGACGACTACTCTAGAAATGCTTGGGTCTCTAAACGGGGTCCATCTGT 383  
|||||  
Oy 3225 TACTGAAAGAGGTGCTGACGTCATCCTTTGGCAAGAGAACAGAGGTGAAATAGA 3284  
|||||

Db	384	TACTGAAAGAGGTGCTGACGTGCATCCTTTGCCAAGAGACAGGAGGTGAAATAGA	443
Oy	3285	AAATAATGCCATGGTATTATCGGCCCTGTCTCCAGAAATCTACTGCTTAAACCCAGCACAG	3344
Db	444		
Db	444	AAATAATGCCATGGTATTATCGGCCCTGTCTCCAGAAATCTACTGCTTAAACCCAGCACAG	503
Oy	3345	GGGAAACCCATAGAACTCTCAGGAGAGCCCTAGACCCACTTTTTCATGGATCCAGACTT	3404
Db	504		
Oy	3405	GGCATATGGAACCTTATACAGGAAGCTGTGGTTCATGTAAATGCAGCGAGTGTGCTGGCAGAA	3464
Db	564	GGCATATGGAACCTTATACAGGAAGCTGTGGTTCATGTAAATGCAGCGAGTGTGCTGGCAGAA	623
Oy	3465	GTATTTTGAAGCTGTACAGCTGAGCTCTCAGCAGCGCATTCATGTTGACCTTTTGGACTT	3524
Db	624	GTATTTTGAAGCTGTACAGCTGAGCTCTCAGCAGCGCATTCATGTTGACCTTTTGGACTT	683
Oy	3525	GGAAAGTGGAGAAATATCTTTGGCCCTCTTTGCAAAATCTCTGTGCAATACTGTGTATCCCAT	3584
Db	684	GGAAAGTGGAGAAATATCTTTGGCCCTCTTTGCAAAATCTCTGTGCAATACTGTGTATCCCAT	743
Oy	3585	TATTCCTTTGCAACCTCAAAAGATAAACAGTGAG	3618
Db	744	TATTCCTTTGCAACCTCAAAAGATAAACAGTAG	777

Search completed: September 27, 2003, 01:57:59  
Job time : 840.105 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 27, 2003, 01:01:23 ; Search time 221.076 Seconds  
(Without alignments)  
10391.910 Million cell updates/sec

Title: US-09-724-126a-18  
Perfect score: 5205  
Sequence: 1 atggcgagcagagagagctg.....tcaactgcagcttactctga 5205

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 569978 segs, 220691566 residues  
Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: /cgn2\_6/ptodata/2/ina/5a\_COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5b\_COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6a\_COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6b\_COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4085.6	78.5	6395	2	US-08-982-956-1
2	4085.6	78.5	6395	3	US-09-228-317-1
3	996.2	19.1	1001	2	US-08-982-956-2
4	996.2	19.1	1001	3	US-09-228-317-2
5	57.4	1.1	7218	1	US-08-232-463-14
6	52.4	1.0	7218	1	US-08-232-463-14
7	44.4	0.9	7333	3	US-09-138-024-21
8	44.4	0.9	7333	4	US-09-404-066-21
9	44.4	0.9	7333	4	US-09-573-322-21
10	40.6	0.8	41708	4	US-09-470-512A-3
11	39.8	0.8	2861	1	US-08-299-953-1
12	39.8	0.8	2861	1	US-08-459-415-1
13	39.8	0.8	2861	4	US-09-066-687-1
14	39.8	0.8	2861	5	PCT-US95-11231-1
15	39.8	0.8	2861	1	US-08-299-953-2
16	39.8	0.8	3881	1	US-08-459-415-2
17	39.8	0.8	3881	4	US-09-066-687-2
18	39.8	0.8	3881	5	PCT-US95-11231-2
19	39.8	0.8	4376	1	US-08-119-125A-1
20	39.4	0.8	751	4	US-09-687-698-11
21	38.4	0.7	5703	4	US-09-280-590A-36
22	38	0.7	1785	4	US-09-601-198-156
23	37.8	0.7	1664976	4	US-08-916-421B-1
24	37.8	0.7	1664976	4	US-08-916-421B-1
25	37	0.7	1113	1	US-08-341-538A-1
26	37	0.7	1113	2	US-08-725-518-1
27	37	0.7	1618	3	US-08-889-108-1

C 28	37	0.7	1618	3	US-08-889-108-3	Sequence 3, Appl1
C 29	37	0.7	1618	3	US-08-120-601B-1	Sequence 1, Appl1
C 30	37	0.7	1618	3	US-08-120-601B-3	Sequence 3, Appl1
C 31	37	0.7	1618	5	PCT-US94-10358-1	Sequence 1, Appl1
C 32	37	0.7	1618	5	PCT-US94-10358-3	Sequence 3, Appl1
C 33	37	0.7	4875	1	US-08-460-739-1	Sequence 1, Appl1
C 34	36.8	0.7	1020	4	US-09-690-454-35	Sequence 35, Appl1
C 35	36.2	0.7	9636	1	US-08-323-170B-1	Sequence 1, Appl1
C 36	36.2	0.7	9636	4	US-08-954-441-1	Sequence 1, Appl1
C 37	36.2	0.7	19124	2	US-08-487-826B-13	Sequence 13, Appl1
C 38	36	0.7	99500	4	US-09-798-096-10	Sequence 10, Appl1
C 39	35.8	0.7	3349	4	US-09-336-447A-2	Sequence 2, Appl1
C 40	35.8	0.7	203001	4	US-09-734-674-3	Sequence 3, Appl1
C 41	35.6	0.7	2319	4	US-09-107-532A-2070	Sequence 2070, AP
C 42	35.6	0.7	5252	4	US-09-340-620A-31	Sequence 51, Appl1
C 43	35.4	0.7	1449	4	US-09-134-1001C-1472	Sequence 1472, Ap
C 44	35.4	0.7	2135	3	US-08-430-286A-1	Sequence 1, Appl1
C 45	35.2	0.7	289	3	US-09-007-005-17	Sequence 17, Appl1

# ALIGNMENTS

```

RESULT 1
US-08-982-956-1
; Sequence 1, Application US/08982956
; Patent No. 5861312
; GENERAL INFORMATION:
; APPLICANT: Varshavsky, Alexander
; TITLE OF INVENTION: NUCLEIC ACID ENCODING MAMMALIAN UBRI
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kevin M. Farrell, P.C.
; STREET: P.O. Box 999
; CITY: York Harbor
; STATE: ME
; COUNTRY: US
; ZIP: 03911
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/982,956
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Farrell, Kevin M.
; REGISTRATION NUMBER: 35,505
; REFERENCE/DOCKET NUMBER: CIT-2001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (207) 363-0558
; TELEFAX: (207) 363-0528
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6395 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 115..5385
; US-08-982-956-1

Query Match      78.5%  Score 4085.6; DB 2; Length 6395;
Best Local Similarity 86.6%  Pred. No. 0;
Matches 4567; Conservative 9; Mismatches 629; Indels 69; Gaps 4;

1 ATGGCGAGCAGAGAGCTGAGAGTACTGAGAGATGGAATCAGCCGAGTTACCCAG 60
|||||

```





Db 4495 GATCTGTCTCCAGGCGCCGCTTGTCTGAGGGTGAAGAGGATAGTGAGGAGGCTCGCTGT 4554  
Qy 4372 GCATCTCTCTTCTTTGTCAGAAATTTCTCAATATACAACTGGCTCCATTTGGGTGTGATATT 4431  
Db 4555 GCATCTGTCTTCTTTGTCAGAAATTTCTCAATATACAACTGGCTCCATTTGGGTGTGATATT 4614  
Qy 4432 CCTGGTGTATTTGTCAGAAATTTCTCAATATACAACTGGCTCCATTTGGGTGTGATATT 4491  
Db 4615 CCCGGTGTATTTGTCAGAAATTTCTCAATATACAACTGGCTCCATTTGGGTGTGATATT 4674  
Qy 4492 GCATCTGTCTTCTTTGTCAGAAATTTCTCAATATACAACTGGCTCCATTTGGGTGTGATATT 4551  
Db 4675 GCATCTGTCTTCTTTGTCAGAAATTTCTCAATATACAACTGGCTCCATTTGGGTGTGATATT 4734  
Qy 4552 GCAGAGGAGAGTACAGTGCCTCTGTAGCTATCTATCTTTACCTACAAATTTGTCCTG 4611  
Db 4735 GCTGAAGGAGAAATTCAGTGCCTCTGTAGCTATCTATCTTTACCTACAAATTTGTCCTG 4794  
Qy 4612 CTCCTCCAGGAATATTGGGATCTGTAAAGCCCTTGTCCAGAGGCGGTGTGCAGATCCT 4671  
Db 4795 CTTTCCAGGAATATTGGGATCTGTAAAGCCCTTGTCCAGAGGCGGTGTGCAGATCCT 4854  
Qy 4672 GCCTTACTAACTGTTTGAAGCAAAACACCGTGGTCACTACCTAGCAAAAGAAAT 4731  
Db 4855 GCCTTACTAACTGTTTGAAGCAAAACACCGTGGTCACTACCTAGCAAAAGAAAT 4914  
Qy 4732 AGTTGTAGAGCTTCTGTAGCTATCTATCTTTACCTACAAATTTGTCCTGAGG 4791  
Db 4915 AGTTGTAGAGCTTCTGTAGCTATCTATCTTTACCTACAAATTTGTCCTGAGG 4974  
Qy 4792 TGCCACGGTGTGAGATGATGAGGAAAGCATCTGTCTCTCAAGCTTCTCATTTACG 4851  
Db 4975 TGTCACGGTGTGAGATGATGAGGAAAGCATCTGTCTCTCAAGCTTCTCATTTACG 5034  
Qy 4852 ATACTATCTCTCAAGCAATTTGTCGACAGAAATTTGAACGGGAGAGGTTGAGCT 4911  
Db 5035 ATCTGTCTCTCAAGCAATTTGTCGACAGAAATTTGAACGGGAGAGGTTGAGCT 5094  
Qy 4912 TGCATTTTTCAGGCACTTCACTGTGGAGCGGAGTCTGCATTTTCTCAAAATCAGAGAA 4971  
Db 5095 TGCATTTTTCAGGCACTTCACTGTGGAGCGGAGTCTGCATTTTCTCAAAATCAGAGAA 5154  
Qy 4972 TGCCAGGTGTGTCGTTGTAAGGTAAGCAAGGCTGTCTCTCAAGCTTCTCATTTACG 5031  
Db 5155 TGCAGGTGTGTCGTTGTAAGGTAAGCAAGGCTGTCTCTCAAGCTTCTCATTTACG 5214  
Qy 5032 GATGAATATGAGAAACAGACCTTGGCTTGAAGGGGCAACCCCTTCAATTTATCTCGT 5091  
Db 5215 GATGAATATGAGAAACAGACCTTGGCTTGAAGGGGCAACCCCTTCAATTTATCTCGG 5274  
Qy 5092 GAGCGGTATCGAAGCTCCATTTGGTCTGGCAACACACTGCATTTATAGAGAGATTTGCT 5151  
Db 5275 GAGCGGTATCGAAGCTCCATTTGGTCTGGCAACACACTGCATTTATAGAGAGATTTGCT 5334  
Qy 5152 AGGAGCCAGAGACTTAATCAGATGTTATTTGGATTCAACTGGCAGTTACTGTGA 5205  
Db 5335 CGGAGCCAGGAGACTTAATCAGATGTTATTTGGATTCAACTGGCAGTTACTGTGA 5388

## RESULT 2

US-09-228-317-1  
; Sequence 1, Application US/09228317  
; Patent No. 6159732  
; GENERAL INFORMATION:  
; APPLICANT: Varshavsky, Alexander  
; APPLICANT: Kwon, Yong Tae  
; TITLE OF INVENTION: NUCLEIC ACID ENCODING MAMMALIAN UBRI  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Kevin M. Farrell, P.C.  
; STREET: P.O. Box 999  
; CITY: York Harbor  
; STATE: ME

COUNTRY: US  
ZIP: 03911  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/228,317  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Farrell, Kevin M.  
REGISTRATION NUMBER: 35,505  
REFERENCE/DOCKET NUMBER: CIT-2001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (207) 363-0558  
TELEFAX: (207) 363-0528  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 6395 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 115..5385  
US-09-228-317-1

Query Match 78.5%; Score 4085.6; DB 3; Length 6395;  
Best Local Similarity 86.6%; Pred. No. 0;  
Matches 4567; Conservative 9; Mismatches 629; Indels 69; Gaps 4;

Qy 1 ATGGCGGAGAGAGGCTGAGAGTACTGAGAGGATGAAATCAATCAAGCGGAGTTACCCAG 60  
Db 115 ATGGCGGAGAGAGAGTGGACGCGCGAGAGGATGAGCTCAGCCGAGCCCTCCCTG 174  
Qy 61 ACCCTTCACGCTCTGGCATCTTGGTGGGATCAGCAAGTTGATTTTATCTGCTTCTTG 120  
Db 175 GCCCGCAGCGCGCGCATCGTGGTGGGATCAGCAAGTTGATTTCTATCTGCTTCTTA 234  
Qy 121 CATCATTTGGCAAAATTTGTCGAGAAATTTACTTTGCTGAAATGGACCCAGACTTGGAA 180  
Db 235 CATCATTTGGCAAAATTTAGTGGCAGAAATTTATTTGCTGAGATGGACCCAGATTGGAA 294  
Qy 181 AAGCAGGAGAAAGTGTACAAATGTCAATATTCACCTCCACTGGAATGGTACTTATTTGA 240  
Db 295 AAGCAGGAGAGAGTGTACAGATGTCAATCTCACTCCTTTGGAGTGGTACTTATTTGA 354  
Qy 241 GAAGATCCAGATATTTGCTTAGAGAAATTTGAAGCAGAGTGGAGCATTTTCAGCTTTGGG 300  
Db 355 GAGGATCCCGATATTTGCTTAGAGAAATTTAAACACAGTGGAGCGGTCCAGTTGTGGG 414  
Qy 301 AGGGTTTTCAAAAGTGAGAGACAACCTATTTCTTGGAGGATTTGTGCAATTTGATCCAA 360  
Db 415 AAGGTTTTCAAAAGTGAGAGACAACATATTTCTGTAGGATTTGTGCAATTTGATCCAA 474  
Qy 361 TGTGTACTCTGTATGGACTGCTTCCAGGACAGTGTTCATATAAATCATCTGTACAGATG 420  
Db 475 TGTGTCTCTGTATGGACTGCTTCCAAAGTAGTGTTCATATAAATCATCTGTACAGATG 534  
Qy 421 CATACTTCTACTGGAGGAGGTTTCTGTACTGTGGAGACACAGAGCATGGAACATGGC 480  
Db 535 CATACTTCTACTGGAGGAGGTTTCTGTACTGTGGAGACACAGAGCGTGGAAACATGGC 594  
Qy 481 CCTTTTGTGTAATCATGAACCTGGAAGCAGAGTACTATAAAGAGAAATTCACGCTGT 540  
Db 595 CCTTTTGTGTAATCATGAACCTGGAAGCAGAGTACTATAAAGAGAGCTTACATTCG 654  
Qy 541 CGGTTGAATGAAGAGGTAATTTGTCAGCCAGCAAAATATTTCTTCTCAGTGATATAAT 600  
Db 655 CCATTGAATGAAGAGGTAATTTGTCAGCCAGGAGAAATATTTCTCTCGGTGATATAAT 714



Db	2875	GGCTTGCTGGAAGAGAACGACGCTTCAGAAAGCTCTGAGAGAGAGTGGCTTTTGAC	2934	Db	3955	AGTTTGGAGTTCAGTCTTCGGTGAATATTCAATAGTATCAAGGAATGCTCATTTCTC	4014
Qy	2776	TTTATATCAATAGGCTTCAAGATTTGGGAATGTCAGCCAT-----GAATATACAAATG	2826	Qy	3844	TTTGGCCACAACAATTTATAGAAATGAGTTGAAAGTCCACCTGATGAAAGGGATTCCTCGA	3903
Db	2935	TTTATACCAATAAGCTTCAAGATTTGGGAATGTCAGCCATGAAATGCTCAGAAATATACAAATG	2994	Db	4015	TTTGGCCACAACAATTTATAGAAATGAGTTGAAAGTCCACCTGATGAACATGAGACCCACGA	4074
Qy	2827	CTTTTGGAAAACCTCAAGGAATTTCCCCAGTTAGNAGCCAGAGGACATGATACGTGG	2886	Qy	3904	GTCCCATCTGACCTGGAGGACCTCGCTTTTCACATATCCAGCAATTTGAAATTCATTG	3963
Db	2995	CTCTTGGAAAGACTCAAGGAATTTCCCCAAATAGAAAGCCAGAGGACATGATACATGG	3054	Db	4075	GTGCCCACATGACCTGGAGCAGTGTGGTTCCACCATCAGGCAATTCGAAACCTGTG	4134
Qy	2887	ATACTTCAGATGTTTGACACAGTGAAGGATTAAGAGAAAATCTTTGTTAATTTGATGCA	2946	Qy	3964	GGAGATGAAGGAAAACCTCTGTTTGGAGCACCTTCAAAATAGGAGCAGATATGCTGTGAAA	4023
Db	3055	ATACTCCAGATGTTTGACACAGTGAAGGATTAAGAGAAAATCTTTGTTAATTTGATGCA	3114	Db	4135	GGAGATGAAGGAAAACCTCTTATTGGAGCACCTTCAAAATAGAGCAGATACGGGTCTGAAG	4194
Qy	2947	ACCACATCAGATCGGAATCTATTAAAGAAATGATGAGATTAATCTCATGATAAAGAAAAGCA	3006	Qy	4024	GCATTAATGCAATTTGACAGTTGCACAGAGGATTTACGTCTCTCAGGTCTCTGATACAGAAA	4083
Db	3115	ACCACATCAGGACTGGAGTGAATTAAGAGTGAAGGATTAATCTCATGATAAAGAAAAGCA	3174	Db	4195	GGCCTAAATGCAGTTTGCAGTTGCACAGAGGGCTACCTGCGCTCAGGTCTCTGATACACAA	4254
Qy	3007	GAACGAAAACAAAGCTGAGCTGCTAGGCTACATCCGAGAGATCATGGCTCAGATG	3066	Qy	4084	CATCTGGTTCGCTCTCTATCAGTTGTTCTTCCCTAACATAAATCAGAAATACACCATGC	4143
Db	3175	GAACGGAAGAAAGCTTGAGCCGCTAGGCTTCATGCCAGAGATCATGGCCAGATG	3234	Db	4255	CATCTGGCTCGGCTCCTGTCAAGTTATTCCTTAACCTGCAATCAGAAATACACGAGC	4314
Qy	3067	TCTGCCCTTACAGAAAACCTTCATTGAAACTCATAACTCATGATGACAAATACATCAGAA	3126	Qy	4144	CTTCTGCTATAGATCTGTTTTCATGTTTGGTGGGTGCTGTGTAGCATTTCCCATCCTTG	4203
Db	3235	TCTGCCCTTACAGAAAACCTTCATTGAAACTCATAACTCATGATGACAAATACATCAGAA	3294	Db	4315	CTTCTGCTGTGATCTCTTCCATGTTTGGTGGGCGCAGTCTTAGGTTTCCCATCCTTG	4374
Qy	3127	ATGCGTGGGAAGAGATTCATATGAGGAAGAGACACCCAGCAGTCAAGTCACTAC	3186	Qy	4204	TAATGGGATGACCTGTTGATCTGACGCTTCTTTCAGTTAGTTTCTTCTATACACCTT	4263
Db	3295	GTAACAGGGAAGAGAGATTCATATGAGGAAGAGACACCCAGCAGTCAAGTCACTAC	3354	Db	4375	TATTGGGATGACACCTGGATCTGACGCGTGGCCACTTAGTTTCTTATATAACCACTC	4434
Qy	3187	TCTAGAATGCTTTGGTTCCTAAACGGGTCCATCTGTTACTGAAAAGGAGGTGCTGAC	3246	Qy	4264	TATCTCTTCCATTTGATACCATGGCACATGCTTTCAGATACTACTTACAGTAGACA--	4321
Db	3355	TCTAGAATGCTTTGGTTCCTAAACGGGTCCATCTGTTACTGAAAAGGAGGTGCTGAC	3414	Db	4435	TATCTCTTCCATTTGATACCATGGCGCACATGCTTTCAGATACTCTTACAACAGATACA	4494
Qy	3247	TGCATCTCTTCCAGAAAGACAGGAGGTGAAAATAGAAAATATGCCATGTTATATCG	3306	Qy	4322	-----CAGGCGTACCCCTGCTCAGGTTTCAAGAGACAGTGAAGAGGCTCATTC	4371
Db	3415	TGCATCTCTTCCAGAAAGACAGGAGGTGAAAATAGAAAATATGCCATGTTATATCG	3474	Db	4495	GATCTGCTCAGGCGCGCGCTGCTGAGGGTGAAGAGGATAGTGAAGAGGCTCGCTGT	4554
Qy	3307	GCCTGTGTCAGAAATCTACGTGCTTAAACGAGCAGAGGGAACCCATAGAACTCTCA	3366	Qy	4372	GCATCTCTTCTTTGTCAGAAAATTTCTCAATATACAAAGTGGTCCATTTGGGTGTGATTT	4431
Db	3475	GCATGTGTGAGAAATCACCGCCCTAACCCAGCAGACAGAGGAGCCCTGTGACCACTTA	3534	Db	4555	GCATCTGCTTTCTTTTGGAAAGTGTGCGACACACAGAGCGGCTCCTACGCGGTGCT	4614
Qy	3367	GGAGAGCCCTAGACCCACTTTTCATGATCCAGACTTGGCATATGGAACCTATACAGGA	3426	Qy	4432	CCTGCTGCTATTTGTTGGGTCTCAGTGAAGATGGCATCACCCCTTCTCTCCCTGCT	4491
Db	3535	GGGAAACACCTGGACCCCTCTTTTATGGATCCAGACTTGGCACATGGAACCTATACAGGA	3594	Db	4615	CCCGCTGTGATCTGTTGGGTCTCCCTGAGGAAGGCGCATCACCCCTTACCTCCGCTGT	4674
Qy	3427	AGCTGTGTCATGTAATGCACGACGTGTGCTGGCAGAAAGTATTTTGAAGCTGTACAGCTG	3486	Qy	4492	GCATTTGTTTTCACCTATTTTACTTGGGGTAACTCCGCGCTGAGGAACCTGATACCAATCT	4551
Db	3595	AGCTGTGTCATGTAATGCATGCAGTGTGCTGGCAGAAAGTATTTTGAAGCTGTGAGCTG	3654	Db	4675	GCATGCTTTTCCACTATTTTACTTGGAGTAGCTCCGCGCTGAAGAACTGTTTGCAATCT	4734
Qy	3487	AGCTCTCAGCAGCGGATTCATGTTGACCTTTTGGACTTGGAAAGTGGAGAAATATCTTGC	3546	Qy	4552	GCAGAAAGGAGATGACAGTGCATCTGTAGTATCTATCTTTACCTACAAATTTGTCCTG	4611
Db	3655	AGCTCGCAGCAGCGCATTCAGTGTAGCCCTGTTGACCTGGAGCGCGGAGTACCTATGC	3714	Db	4735	GCTGAAGGAGAAATTCAGTGCACCTGTAGTATCTATCTTTTACCCACAAATTTGTCCTG	4794
Qy	3547	CCTCTTTCGCAATCTCTGTGCAATCTGTGATCCCATTTATTCCTTTCGCAACCTCAAAAG	3606	Qy	4612	CTCTTCCAGGAATATTGGGATACGTGAAGCCCTTCTCCAGAGGCGGTGTCAGATCTCT	4671
Db	3715	CCGCTCTGCAAGTCTCTCTGCAACACTGTCTATCCCATCTATCCCTTTTCGAGCCGAGAG	3774	Db	4795	CTTTTCCAGGAATATTGGGATACCATAGGCCCTTACTACAGAGGTGGTGGAGATCTCT	4854
Qy	3607	ATAACAGTGAATGACATGCTCTTCTCAACTTTTGAACCTGGCAGGCTGGATACAG	3666	Qy	4672	GCCTTACTAACTTTTGAAGCAAAAACACCGCTGCTGAGTACCCCTAGAAAAAGAAAT	4731
Db	3775	ATCAACAGTGAATGCGGAGCTCTTGCTCAACTTTTGACCTTTGGCCCGGTGGATACAG	3834	Db	4855	GCCTTACTCAAGTCTTTGAAGCAGAAAAGTGTGTGTGTCAGGTACCCCTAGAAAAAGAAAT	4914
Qy	3667	ACTGTTTGGCCAGAAATACAGGTTATTAATAAGACATGCTAAAGGAGAAAAC--CA	3723	Qy	4732	AGTTTGAATAGAGTCTCTGATGACTATAGTGTGCTCTCTCTCTCTCTCTCTCTCTCTCT	4791
Db	3835	ACTGTCTTGCAGAAATATCGGGTTATAATAAAGCATGCTAAAGGAGAAAGCCCAAGCA	3894	Db	4915	AGTTTGAATAGAGTCTCTGATGAGGACTACAGTGTCTTCTAAATCAGGCTTCTCACTTAGG	4974
Qy	3724	ATTCTATTTTCTTAATCAAGGAATGGGAGATTTCTACTTTGGAGTTCCATTCATCTG	3783	Qy	4792	TGCCCACGCTCTGCAGATGATGAGCCAAAGCATCTCTCTCTCTCTCTCTCTCTCTCTCT	4851
Db	3895	GTTCCTGTCTTTTAAATCAAGGAATGGGGATCAACTTTTGTGTTTCAATTCATCTCTG	3954	Db	4975	TGTCACGCTCTGCAGATGATGAGCGAAAAGCATCTCTCTCTCTCTCTCTCTCTCTCTCT	5034
Qy	3784	AGTTTGGCGTGTAGTCTTCGATTAATAATTTCAATACCATCAAGGAATGGTTATTCTC	3843	Qy	4852	ATACTATGTTCTCAGAACATTTGCTGCCAGGAATTTGTGAACGGGAAGAGGTTGGAGCT	4911
				Db	5035	ATCTGTGTTCTCAGAACATCTGTTGCCAGAAATAGTGAATGGGAGAGGTTGGAGCG	5094



QY 4912 TGCATTTTTCAGCAGCTTCACTGTGAGCGGAGCTGCATTTTCTAAATACAGAA 4971  
DB 5095 TCGGTTTTTCATGCGCTTCATGTGGTGTGAGTCTGCATTTTCTAAATACGAGAA 5154  
QY 4972 TGGCAGTGGTCTGTTTGAAGGTAAAGCCAGAGGCTGTGCCATATCCAGCTCCTTACTTG 5031  
DB 5155 TGCAGGTGGTCTGTTGGAAGGAAGAACAGAGGCTGTGCCATCCAGCCCTTACTTG 5214  
QY 5032 GATGAATATGAGAAACAGACCCCTGGCCTGAAGAGGGCAACCCCTTCATTTATCTCGT 5091  
DB 5215 GATGAATATGAGAAACAGACCCAGGCTAAAGAGAGAAACCCACTTCATTTATCTCGG 5274  
QY 5092 GAGCGGTATCGAAGCTCCATTTGTTCTGGAACAACTGATATATAGAGAGATTGCT 5151  
DB 5275 GAGCGGTATCGAAGCTCCATTTGTTCTGGAACAACTGATATATAGAGAGATTGCT 5334  
QY 5152 AGGAGCCAAAGAGACTAATCAGATGTTATTTGGATTCAACTGCGACTTACTGTGA 5205  
DB 5335 CGGAGCCAGGAGACTAATCAGATGCTATTTGGATTCAACTGCGACTTACTGTGA 5388

US-08-982-956-2  
; Sequence 2, Application US/08982956  
; Patent No. 5861312

GENERAL INFORMATION:  
; APPLICANT: Varshavsky, Alexander  
; APPLICANT: Kwon, Yong Tae

TITLE OF INVENTION: NUCLEIC ACID ENCODING MAMMALIAN UBRI  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Kevin M. Farrell, P. C.  
; STREET: P.O. Box 999  
; CITY: York Harbor  
; STATE: ME  
; COUNTRY: US  
; ZIP: 03911

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/982,956  
; FILING DATE:

CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Farrell, Kevin M.  
; REGISTRATION NUMBER: 35,505  
; REFERENCE/DOCKET NUMBER: CIT-2001

TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (207) 363-0558  
; TELEFAX: (207) 363-0528  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1001 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..999

US-08-982-956-2

Query Match 19.1%; Score 996.2; DB 2; Length 1001;  
Best Local Similarity 99.7%; Pred. No. 3.3e-276;  
Matches 998; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2059 ATGGATCCCAATAGTTCTTGTACTGTGTACTTTCAGAGGTATGAACCTGCCAGGCTTTT 2118  
DB 1 ATGGATCCCAACAGTTCTTGTACTGTGTACTTTCAGAGGTATGAACCTGCCAGGCTTTT 60

QY 2119 AACAAGACCATACTACAAAAGACCAGGATTTGATTAAACAATATATACACTAATAGAA 2178  
DB 61 AACAAGACCATACTACAAAAGACCAGGATTTGATTAAACAATATATACACTAATAGAA 120  
QY 2179 GAAATGCTTCAGGCTCCTCATCTATATTTGTTGGTGAGCGTTATGTACCTTGGAGTGGGAAT 2238  
DB 121 GAAATGCTTCAGGCTCCTCATCTATATTTGTTGGTGAGCGTTATGTACCTTGGAGTGGGAAT 180  
QY 2239 GTGACCAAAAGAGAGGTCACAATGAGAGAAATCATTTCACTTGTCTTTGATTGAACCCATG 2298  
DB 181 GTGACCAAAAGAGAGGTCACAATGAGAGAAATCATTTCACTTGTCTTTGATTGAACCCATG 240  
QY 2299 CCACACAGTGCCTATTCGCAAAAATTTACCTGAGAAATGAAAATATATGAACTGGCTTAGAG 2358  
DB 241 CCACACAGTGCCTATTCGCAAAAATTTACCTGAGAAATGAAAATATATGAACTGGCTTAGAG 300  
QY 2359 AATGTCTATAAACAAGTGGCCACATTTAAGAAACCAGGTGTATCAGGCCCATGGAGTTTAT 2418  
DB 301 AATGTCTATAAACAAGTGGCCACATTTAAGAAACCAGGTGTATCAGGCCCATGGAGTTTAT 360  
QY 2419 GAACTAAAAGATGAATCACTGAAAGACTTCAATATGTACTTTTATCATTTACTCCAAAACC 2478  
DB 361 GAACTAAAAGATGAATCACTGAAAGACTTCAATATGTACTTTTATCATTTACTCCAAAACC 420  
QY 2479 CAGCATAGCAAGGCTGAACATATGCAAGAGAAAAGAGAGAAAACAAGAAAACAAGATGAA 2538  
DB 421 CAGCATAGCAAGGCTGAACATATGCAAGAGAAAAGAGAGAAAACAAGAAAACAAGATGAA 480  
QY 2539 GCATTCGCCGCCACCACCCTCCTGATTTCTGCCCTGCTTTTCAGCAAGTGAATTAACCTT 2598  
DB 481 GCATTCGCCGCCACCACCCTCCTGATTTCTGCCCTGCTTTTCAGCAAGTGAATTAACCTT 540  
QY 2599 CTCAACTGTGATATCATGTGTACATTTCTCAGGACCGCTATTTGAGCGGGCAATAGACACA 2658  
DB 541 CTCAACTGTGATATCATGTGTACATTTCTCAGGACCGCTATTTGAGCGGGCAATAGACACA 600  
QY 2659 GATTTCAACTTGTGACCGGAGGATGCTCCAAATGGCTTTTTCATATTTCTGGCATTTGGGT 2718  
DB 601 GATTTCAACTTGTGACCGGAGGATGCTCCAAATGGCTTTTTCATATTTCTGGCATTTGGGT 660  
QY 2719 TTACTAGAGAGAGCAACAGCTTCAAAAAGCTCTCTGAAGAGAGTGAACATTTGACTTT 2778  
DB 661 TTACTAGAGAGAGCAACAGCTTCAAAAAGCTCTCTGAAGAGAGTGAACATTTGACTTT 720  
QY 2779 TATCATAAGCTTCAAGATTGGGAAGTTTCAGCCATGAATATACAAATGCTTTTGGAAAAA 2838  
DB 721 TATCATAAGCTTCAAGATTGGGAAGTTTCAGCCATGAATATACAAATGCTTTTGGAAAAA 780  
QY 2839 CTCAAAAGGAATTTCCCAGTTTAGAAGGCCAGAAGSACATGATAACGTGGATACTTTCAGATG 2898  
DB 781 CTCAAAAGGAATTTCCCAGTTTAGAAGGCCAGAAGSACATGATAACGTGGATACTTTCAGATG 840  
QY 2899 TTTGACACAGTGAAGCGATTAAAGAAAAATCTTTTAAATTTAGCAACACATCAGGA 2958  
DB 841 TTTGACACAGTGAAGCGATTAAAGAAAAATCTTTTAAATTTAGCAACACATCAGGA 900  
QY 2959 TCGGAATCTATTAAAGAAATGATGAGATTACTCATGATAAAGAAAAGCAGAACGAAAAAGA 3018  
DB 901 TCGGAATCTATTAAAGAAATGATGAGATTACTCATGATAAAGAAAAGCAGAACGAAAAAGA 960  
QY 3019 AAAGCTGAAGCTGCTAGGCTTACATCGCCAGAAGATCATGGC 3059  
DB 961 AAAGCTGAAGCTGCTAGGCTTACATCGCCAGAAGATCATGGC 1001

RESULT 4  
US-09-228-317-2  
; Sequence 2, Application US/09228317  
; Patent No. 6159732  
; GENERAL INFORMATION:  
; APPLICANT: Varshavsky, Alexander  
; APPLICANT: Kwon, Yong Tae

;; TITLE OF INVENTION: NUCLEIC ACID ENCODING MAMMALIAN UBRI  
;; NUMBER OF SEQUENCES: 2  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Kevin M. Farrell, P.C.  
;; STREET: P.O. Box 999  
;; CITY: York Harbor  
;; STATE: ME  
;; COUNTRY: US  
;; ZIP: 03911  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: Patent In Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/09/228,317  
;; FILING DATE:  
;; CLASSIFICATION:  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Farrell, Kevin M.  
;; REGISTRATION NUMBER: 35,505  
;; REFERENCE/DOCKET NUMBER: CIT-2001  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (207) 363-0558  
;; TELEFAX: (207) 363-0528  
;; INFORMATION FOR SEQ ID NO: 2:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 1001 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: cDNA  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: 1..999  
;; US-09-228-317-2

Query Match 19.18; Score 996.2; DB 3; Length 1001;  
Best Local Similarity 99.7%; Pred. No. 3.3e-276;  
Matches 998; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY	2059	ATGGATCCCAATAGTCTTGTTACTGTCTACTCAGAGGTATGAACCTGCGAGGCTTTT	2118
DB	1	ATGGATCCCAACAGTCTTGTTACTGTCTACTCAGAGGTATGAACCTGCGAGGCTTTT	60
QY	2119	AACAAGACCATATCTACAAAAGACGAGGATTTGATTAAACAATAATACTACTAATAGAA	2178
DB	61	AACAAGACCATATCTACAAAAGACGAGGATTTGATTAAACAATAATACTACTAATAGAA	120
QY	2179	GAATGCTTCAGGTCCTCATCTATATTGTTGGGTGAGCGTTATGTTACCTGGAGTGGAAAT	2238
DB	121	GAATGCTTCAGGTCCTCATCTATATTGTTGGGTGAGCGTTATGTTACCTGGAGTGGAAAT	180
QY	2239	GTGACCAAAAGAGGTCACAATGAGAGAATCATCTTCATTCGCTTCATTCGACCCCATG	2298
DB	181	GTGACCAAAAGAGGTCACAATGAGAGAATCATCTTCATTCGCTTCATTCGACCCCATG	240
QY	2299	CCACAGTGCCATTTGTCACAAAATTTACCTGAGAAATGAAATGAACTGGCTTAGAG	2358
DB	241	CCACAGTGCCATTTGTCACAAAATTTACCTGAGAAATGAAATGAACTGGCTTAGAG	300
QY	2359	AATGTCATAACAAAGTGGCCACATTTTAAAGAACCCAGGTGTATCAGGCCATGGAGTTTAT	2418
DB	301	AATGTCATAACAAAGTGGCCACATTTTAAAGAACCCAGGTGTATCAGGCCATGGAGTTTAT	360
QY	2419	GAACATAAAGATGAATCAGTGAAGACTTCAATATGTACTTTTATCATCTCCAAACCC	2478
DB	361	GAACATAAAGATGAATCAGTGAAGACTTCAATATGTACTTTTATCATCTCCAAACCC	420
QY	2479	CAGCATAGCAAGGCTGAACATATGTCAGAGAAAGAGGAGAAACAAAGAAAGATGAA	2538
DB	421	CAGCATAGCAAGGCTGAACATATGTCAGAGAAAGAGGAGAAACAAAGAAAGATGAA	480

QY	2539	GCATTGCCGCCACCACCCTCCTCAATTTCTGCCCTGCTTTTCAGCAAAAGTGATTACCTT	2598
DB	481	GCATTGCCGCCACCACCCTCCTCAATTTCTGCCCTGCTTTTCAGCAAAAGTGATTACCTT	540
QY	2599	CTCAACTGTGATATCATGATGTACATTTCTCAGAGCCGTTATTTGAGCGGCAATAGACACA	2658
DB	541	CTCAACTGTGATATCATGATGTACATTTCTCAGAGCCGTTATTTGAGCGGCAATAGACACA	600
QY	2659	GATTCCTAACTTGTGGACCGAAGGGATGCTCCAAATGGCTTTTCATATTTCTGGCATTGGGT	2718
DB	601	GATTCCTAACTTGTGGACCGAAGGGATGCTCCAAATGGCTTTTCATATTTCTGGCATTGGGT	660
QY	2719	TTACTAGAAGAGAACACACAGCTTCAAAAGCTCCTGAAGAAGAGTACATTTGACTTT	2778
DB	661	TTACTAGAAGAGAACACACAGCTTCAAAAGCTCCTGAAGAAGAGTACATTTGACTTT	720
QY	2779	TATCATAAAGGCTTCAAGATTTGGGAAGTTTCAGCCATGAATATACAAATGCTTTTGGAAAAA	2838
DB	721	TATCATAAAGGCTTCAAGATTTGGGAAGTTTCAGCCATGAATATACAAATGCTTTTGGAAAAA	780
QY	2839	CTCAAAGGAATTTCCCAAGTTTGAAGGCCAGAGGACATGATACCTGGGATCTTCAGATG	2898
DB	781	CTCAAAGGAATTTCCCAAGTTTGAAGGCCAGAGGACATGATACCTGGGATCTTCAGATG	840
QY	2899	TTTGACACAGTGAAGCGGATTAAGAGAAAAATCTTTTAAATTTGTAGCAACCATCAGGA	2958
DB	841	TTTGACACAGTGAAGCGGATTAAGAGAAAAATCTTTTAAATTTGTAGCAACCATCAGGA	900
QY	2959	TCGGAATCTATTAAAGATGATGAGATTACTCATGATAAAGAAAAAGCAACGAAAAAGA	3018
DB	901	TCGGAATCTATTAAAGATGATGAGATTACTCATGATAAAGAAAAAGCAACGAAAAAGA	960
QY	3019	AAAGCTGAAGCTGTAGGTACATGCCAGAGAGATCATGGC	3059
DB	961	AAAGCTGAAGCTGTAGGTACATGCCAGAGAGATCATGGC	1001

RESULT 5  
US-08-232-463-14/c  
; Sequence 14, Application US/08232463  
; Patent No. 5670367  
; GENERAL INFORMATION:  
; APPLICANT: DORNER, F.  
; APPLICANT: SCHEIFLINGER, F.  
; APPLICANT: FALKNER, F. G.  
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 1800 Diagonal Road, Suite 500  
; CITY: Alexandria  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22313-0299  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/232,463  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/07/935,313  
; FILING DATE:  
; APPLICATION NUMBER: EP 91 114 300.6  
; FILING DATE: 26-AUG-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BENT, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
; TELECOMMUNICATION INFORMATION:





Db 1230 AAAAAATCTTTTATATATTAAATAAATGGAGAGATAAAATTTATATAATTTAAAAAAG 1177

Qy 2987 CTCATGATAAGAAAAGCAGACGAGAAAAGAAAGCTGAAGCTGTAGCTACATCGCC 3046

Db 1170 ATAATAATAAAATCTAGAGTTTATAATAAAAAACTAATATTAATCTCTTTTAACATGTA 1111

Qy 3047 AGAAGATCATGGCTCAGATCTCTGCGCTTACAGAAAAAATCTTCAATGAACTCATAAACTCA 3106

Db 1110 AAATGATTATATTATGATATAATTTTTTTTCAAAACAACCAATAATAAAAAATGATAGG 1051

Qy 3107 TGATGACATACATCAGAAATCGCTGGGAGAA 3141

Db 1050 AGTATTATCATATGTCAGAAATTTATTATAAAGAA 1016

RESULT 12

US-08-459-415-1/c

; Sequence 1, Application US/08459415

; Patent No. 5744334

; GENERAL INFORMATION:

; APPLICANT: Dobres, Michael S. and Mandaci, Sevnur

; TITLE OF INVENTION: A Plant Promoter Useful for Directing the

; FIELD OF INVENTION: Expression of Foreign Proteins to the Plant Epidermis

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5744334ris

; STREET: One Liberty Place 46Ch. Floor

; CITY: Philadelphia

; STATE: PA

; ZIP: 19103

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/459,415

; FILING DATE: 02-JUN-1995

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/299,953

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Beardell, Lori Y.

; REGISTRATION NUMBER: 34,293

; REFERENCE/DOCKET NUMBER: NOVA-0003

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 215-564-8960

; TELEFAX: 215-568-3439

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2861 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

US-08-459-415-1

Query Match 0.8%; Score 39.8; DB 1; Length 2861;

Best Local Similarity 46.5%; Pred. No. 0.68;

	120	conservative	0	mismatches	497	indels	0	gaps	
Qy	2867	AGAGGACATGATAACGTGGATCTACTCAGATCTTTGCACAGTGAACCGATTTAAGAGAA							2926
Db	1290	AAAAAGACGGGGAGTAAATATTTTAGATTTTACACGATTTAAAAAATTTATATCAA							1331
Qy	2927	AATCTTGTTTAAATGTAGCAACCAATCAGGATCGGAATCTATTAGAATGATGAGATT							2986
Db	1230	AAAAATATCTTTATATATTAATAAATGGAGAAGATAAATTTATATAATTTAAAAAAAAG							1171

QY 2987 CTCATGATAAAGAAAGCAGAACGAAAGAAAGCTGAAGCTGCTAGGCTACATCGCC 3046  
 Db 1170 ATAATAATAAATCTAGAGTATATAAATAAATAAATAAATAAATTTCTTTTAAACATGTA 1111  
 QY 3047 AGAAGATCATGCTCAGATGCTGCTTACAGAAAGAACTTCATTAAGAACTCAAACTCA 3106  
 Db 1110 AAATGATTATATATATATAAATTTTTCACAAACCAACATAAATAAATGATAGG 1051  
 QY 3107 TGTATGACAATACATCAAAATGCTGGGAAGAA 3141  
 Db 1050 AGTATTATCATATGTCAGAAATTTATTATAAAGAA 1016

RESULT 13

US-09-066-687-1/c  
 ; Sequence 1, Application US/09066687  
 ; Patent No. 6339185  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Dobres, Michael S. and Mandaci, Sevnur  
 ; TITLE OF INVENTION: A Plant Promoter Useful for Directing the  
 ; TITLE OF INVENTION: Expression of Foreign Proteins to the Plant Epidermis  
 ; NUMBER OF SEQUENCES: 7  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 6339185ris  
 ; STREET: One Liberty Place 46th. Floor  
 ; CITY: Philadelphia  
 ; STATE: PA  
 ; ZIP: 19103  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/066.687  
 ; FILING DATE: Herewith  
 ; CLASSIFICATION:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Beardell, Lori Y.  
 ; REGISTRATION NUMBER: 34,293  
 ; REFERENCE/DOCKET NUMBER: NOVA-0003  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 215-564-8960  
 ; TELEFAX: 215-568-3439  
 ; INFORMATION FOR SEQ ID NO: 1:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 2861 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: DNA (genomic)  
 ; HYPOTHETICAL: NO  
 ; ANTI-SENSE: NO  
 ; US-09-066-687-1

Query Match 0.8%; Score 39.8; DB 4; Length 2861;  
 Best Local Similarity 46.5%; Pred. No. 0.68;  
 Matches 128; Conservative 0; Mismatches 147; Indels 0; Gaps 0;

QY 2867 AGAAGGACATNATACGCTGGATCTTTCAGACAGTGAAGCGATTAAAGAA 2926  
 Db 1290 AAAAAGAACGGAGGAGTTAATTTTACACAGTATTAATAAATAAATAAATATATCA 1231  
 QY 2927 AATCTGTTTAAATGTTAGCAACCATCAGGATCTATTAAAGATGATGATTA 2986  
 Db 1230 AAAATATCTTTATATATAAATCGAGAGATAAATTTATATAATAAATAAAG 1171  
 QY 2987 CTCATGATAAAGAAAGCAGAACGAAAGCTGAAGCTGCTAGGCTACATCGCC 3046  
 Db 1170 ATAATAATAAATCTAGAGTATATAAATAAATAAATAAATAAATTTCTTTTAACTGTA 1111  
 QY 3047 AGAAGATCAAGAAAGCAGAACGAAAGAAAGCTGCTAGGCTACATCGCC 3046  
 Db 1170 ATAATAATAAATCTAGAGTATATAAATAAATAAATAAATAAATTTCTTTTAACTGTA 1111  
 QY 3047 AGAAGATCAAGCTGCTAGGCTTACAGAAAGAACTTCATTAAGAACTCAAACTCA 3106

Db 1110 AAATGATTATATATATATATATTTTTCACAAACCAACATAAATAAATGATAGG 1051  
 QY 3107 TGTATGACAATACATCAGAAATGCTGGGAAGAA 3141  
 Db 1050 AGTATTATCATATGTCAGAAATTTATTATAAAGAA 1016

RESULT 14

PCT-US95-11231-1/c  
 ; Sequence 1, Application PC/TUS9511231  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Dobres, Michael S. and Mandaci, Sevnur  
 ; TITLE OF INVENTION: A Plant Promoter Useful for Directing the Expression  
 ; TITLE OF INVENTION: of Foreign Proteins to the Plant Epidermis  
 ; NUMBER OF SEQUENCES: 4  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & Norris  
 ; STREET: One Liberty Place 46th. Floor  
 ; CITY: Philadelphia  
 ; STATE: PA  
 ; ZIP: 19103  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: PCT/US95/11231  
 ; FILING DATE: Herewith  
 ; CLASSIFICATION:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/299,953  
 ; FILING DATE: September 2, 1994  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Beardell, Lori Y.  
 ; REGISTRATION NUMBER: 34,293  
 ; REFERENCE/DOCKET NUMBER: NOVA-0016  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 215-564-8960  
 ; TELEFAX: 215-568-3439  
 ; INFORMATION FOR SEQ ID NO: 1:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 2861 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: DNA (genomic)  
 ; HYPOTHETICAL: NO  
 ; ANTI-SENSE: NO  
 ; PCT-US95-11231-1

Query Match 0.8%; Score 39.8; DB 5; Length 2861;  
 Best Local Similarity 46.5%; Pred. No. 0.68;  
 Matches 128; Conservative 0; Mismatches 147; Indels 0; Gaps 0;

QY 2867 AGAAGGACATGATAACGCTGGATCTTTCAGACAGTGAAGCGATTAAAGAA 2926  
 Db 1290 AAAAAGAACGGAGGAGTTAATTTTACACAGTATTAATAAATAAATAAATATATCA 1231  
 QY 2927 AATCTGTTTAAATGTTAGCAACCATCAGGATCTATTAAAGATGATGATTA 2986  
 Db 1230 AAAATATCTTTATATATAAATCGAGAGATAAATTTATATAATAAATAAAG 1171  
 QY 2987 CTCATGATAAAGAAAGCAGAACGAAAGCTGAAGCTGCTAGGCTACATCGCC 3046  
 Db 1170 ATAATAATAAATCTAGAGTATATAAATAAATAAATAAATAAATTTCTTTTAACTGTA 1111  
 QY 3047 AGAAGATCAAGCTGCTAGGCTTACAGAAAGAACTTCATTAAGAACTCAAACTCA 3106  
 Db 1110 AAATGATTATATATGATATAAATTTTTCACAAACCAACATAAATAAATGATAGG 1051  
 QY 3107 TGTATGACAATACATCAGAAATGCTGGGAAGAA 3141

Job time : 231.076 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 27, 2003, 09:29:56 ; Search time 814.228 Seconds  
(without alignments)  
15914.219 Million cell updates/sec

Title: US-09-724-126A-18  
Perfect score: 5205  
Sequence: 1 atgaggcagcaggagctgg.....tcaactgacgttactgtga 5205

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1678620 seqs, 1244745471 residues  
Total number of hits satisfying chosen parameters: 3357240

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database : Published Applications\_NA.\*
- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*
  - 2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq.\*
  - 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*
  - 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*
  - 5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*
  - 6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq.\*
  - 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*
  - 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*
  - 9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq.\*
  - 10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq.\*
  - 11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq.\*
  - 12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*
  - 13: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq.\*
  - 14: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq.\*
  - 15: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*
  - 16: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*
  - 17: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	812	15.6	3327	10	US-09-529-063-56 Sequence 56, Appl
2	812	15.6	3327	12	US-10-414-378-56 Sequence 56, Appl
3	812	15.6	3502	10	US-09-529-063-57 Sequence 57, Appl
4	812	15.6	3502	12	US-10-414-378-57 Sequence 57, Appl
5	438.4	8.4	505	11	US-09-918-995-36721 Sequence 36721, A
6	249	4.8	972	9	US-09-822-849A-532 Sequence 532, Appl
7	218.2	4.2	4573	13	US-10-071-766-7 Sequence 7, Appl
8	212.2	4.1	578	13	US-10-027-632-260533 Sequence 260533, A
9	65	1.2	465	11	US-09-918-995-27470 Sequence 27470, A
10	60	1.2	60	12	US-09-908-975-8907 Sequence 8907, Ap
11	53.8	1.0	611	13	US-10-027-632-8115 Sequence 8115, Ap
12	44.2	0.8	535	14	US-10-171-581-278 Sequence 278, Appl
13	44.2	0.8	5398	12	US-10-311-455-251 Sequence 251, Appl
14	44	0.8	1024	10	US-09-895-913A-153 Sequence 153, Appl
15	43.8	0.8	2372	12	US-09-814-353-21640 Sequence 21640, A
16	43.8	0.8	2462	14	US-10-050-704-24 Sequence 24, Appl

c	17	43.8	0.8	6963	12	US-10-311-455-496	Sequence 496, App
	18	43.6	0.8	747	12	US-10-140-472-212	Sequence 212, App
	19	43.6	0.8	747	12	US-10-141-761-212	Sequence 212, App
	20	43.6	0.8	747	12	US-10-142-885-212	Sequence 212, App
	21	43.6	0.8	747	14	US-10-123-155-212	Sequence 212, App
	22	43.6	0.8	747	15	US-10-146-731-212	Sequence 212, App
	23	43.6	0.8	11670	12	US-10-240-452-25	Sequence 25, Appl
	24	43.4	0.8	18988	12	US-10-311-455-674	Sequence 674, Appl
	25	43.4	0.8	18988	12	US-10-240-485-62	Sequence 62, Appl
	26	43.2	0.8	8996	12	US-10-240-453-309	Sequence 309, App
	27	43.2	0.8	8996	14	US-10-239-676-211	Sequence 211, App
	28	42.8	0.8	13574	12	US-10-311-455-1289	Sequence 1289, Ap
	29	42.8	0.8	17137	12	US-10-311-455-163	Sequence 163, App
	30	42.6	0.8	7934	12	US-10-240-453-169	Sequence 169, App
	31	42.6	0.8	7934	14	US-10-239-676-153	Sequence 153, App
	32	42.4	0.8	6265	12	US-10-311-455-1390	Sequence 1390, Ap
	33	42.4	0.8	15861	12	US-10-311-455-497	Sequence 497, App
	34	42.2	0.8	437	10	US-09-796-692-8310	Sequence 8310, Ap
	35	42.2	0.8	437	14	US-10-040-862-8310	Sequence 8310, Ap
	36	42.2	0.8	5557	12	US-10-311-455-1519	Sequence 1519, Ap
	37	42.2	0.8	16994	12	US-10-311-455-961	Sequence 961, App
	38	41.8	0.8	781	13	US-10-027-632-122062	Sequence 122062,
	39	41.4	0.8	5203	12	US-10-311-455-892	Sequence 892, App
	40	41.4	0.8	6080	12	US-10-311-455-1035	Sequence 1035, App
	41	41.4	0.8	10467	12	US-10-240-453-327	Sequence 327, App
	42	41.2	0.8	7657	12	US-10-311-455-1995	Sequence 1995, Ap
	43	41.2	0.8	7657	14	US-10-239-676-185	Sequence 185, App
	44	41.2	0.8	7829	12	US-10-311-455-1078	Sequence 1078, Ap
	45	41	0.8	638	13	US-10-027-632-200125	Sequence 200125,

ALIGNMENTS

RESULT 1  
US-09-529-063-56  
; Sequence 56, Application US/09529063  
; Patent No. US20020102542A1  
; GENERAL INFORMATION:  
; APPLICANT: FUKUSHIMA, DAIKICHI  
; APPLICANT: SHIBAYAMA, SHIRO  
; APPLICANT: TADA, HIDEAKI  
; TITLE OF INVENTION: POLYPEPTIDE, cDNA ENCODING THE POLYPEPTIDE, AND USE OF  
; FILE REFERENCE: 058769  
; CURRENT APPLICATION NUMBER: US/09/529.063  
; CURRENT FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: PCT/JF98/04514  
; PRIOR FILING DATE: 1998-10-06  
; PRIOR APPLICATION NUMBER: JP 9-274674  
; PRIOR FILING DATE: 1997-10-07  
; NUMBER OF SEQ ID NOS: 117  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 56  
; LENGTH: 3327  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-529-063-56

Query Match	15.6%;	Score 812;	DB 10;	Length 3327;
Best Local Similarity	55.3%;	Pred. No. 4.9e-213;		
Matches 1855;	Conservative 0;	Mismatches 1410;	Indels 87;	Gaps 11;
QY	1892	TGGAATACCTTTACGTTGCTGTGTTGGTGGCCAGGTTGTTGCTGAGATGTGGCGAA	1951	
DB	8	TAGAACCCCTCTAGATGCTTCTGTGTGTTGGCCAGTACATGCCGGAATGTGGAGAA	67	
QY	1952	GAATGACGTCTCTTATTAGCCAGGTGTTTATTACCAAGATCTTAAGTCGAGAGAAG	2011	
DB	68	GAATGGTTCTCTCTAGTAACACAGATTTATTACTACCATTAATGTAATGCAGACGTG	127	
QY	2012	AAATGTATGATAAGATATCATCATGCTTCAGATTGGTGCATCTTTAATGATGCCAATA	2071	

Db 128 AGATGTTTGAAGGATGTAGTAATGCTTCAGACAGGTGTCTCCATGATGGATCCAAATC 187

Qy 2072 AGTTCTTGTACTGGTACTCTAGAGGTATGAACCTGGCGAGGCTTTTAAACAAGACCATAT 2131

Db 188 ATTTCTCTGATGATCATGCTCAGCGGCTTTGAACCTTTATCAGATTTTCACTACTCCAGACT 247

Qy 2132 CTACAAAA-----GACCAGGATTTGATTTAAACAATATATATACAC 2170

Db 248 ATGGAAAAAGATTTAGTTCTGAGATTACCCATTAAGGATGTGTTTCAGCAGAACATACTC 307

Qy 2171 TAATAGAAAGAAATGCTTCAGTCTCATCTATATTTGGTGGGTGAGCGTTATGTAAGTGGAG 2230

Db 308 TAATAGAAAGAAATGCTATACCTCATTTATATGCTTTGGAGAGATTTAGTCTCGGAG 367

Qy 2231 TGGGAAATGTGACCAAAAGAGAGGTCAACATGAGAGAAATCAATTCACCTTGCTTTGCAATG 2290

Db 368 TTGGACAGGTAAATGCTACAGATGAATCAAGTAAAGTCTTTACCTGAAGATGAGAACAAAGGAGCTG 427

Qy 2291 AACCCATGCCACAGCTGCCATTCGCCAAATAATTTACCTGAGAAATGAATAATGAACCTG 2350

Db 428 AGCCTATGGCTCATAGTGAATTTGTAAGTCTTTACCTGAAGATGAGAACAAAGGAGACTG 487

Qy 2351 GCTTAGAGAAATGCTATAAAACAAAGTGGCCACATTTAAAGAACACAGGTGTATCAGGCCATG 2410

Db 488 GCATGGAGAGTGAATCGAAGCAGTTGCCATTTCAAGAAACCTGGATTACAGAGCAGG 547

Qy 2411 GAGTTTATGAATTAAGATGAATCACTGAAAGACTTCAATATATGTAATTTTATCATTAAT 2470

Db 548 GCATGTATGAATGAACACAGATGTGCCAAAGATTTCAACTGTATTTCTCATCTTTT 607

Qy 2471 CCAAAACCCAGCATAGCAAGGCTCAACATATGAGAGAAAGAGAGAAACAAACAACA 2530

Db 608 CAAGGGCAGAACAGTCCAAAGCAGAGAAAGCGCAACGGAAATTTGAAAGACAAATAGAG 667

Qy 2531 AAGATGAAGATTCGCCGCCACCCACTCTCTGAATTTCTGCCCTGCTTTTCAGCAAGTGA 2590

Db 668 AAGATCAGACACTCCCACTCCGGTGTGCTCCTCAATCTGCCCTCTGTTTCCAGCCCTGG 727

Qy 2591 TTAACCTTCTCAACTGTGATATCATGATGATACATTTCTCAGGACCGTATTTGAGGGGGCAA 2650

Db 728 TTAACATTTTGCAGTCAGATGTCATGTTGTGCATCATGGGAACAATTTCTGCAATGGGCTG 787

Qy 2651 TAGACACAGATTTCAACTTGTGGACCGAAGGATGCTCCAAATGGCTTTTCAATATCTGG 2710

Db 788 TGGAAACATAATGGATATGCTGTTGTCAGATGCTGCTGCAAGGGTGTGTACATTTAATG 847

Qy 2711 CATTTGGTGTACTAGAGAGAACACAGCTTCAAAAAGCTCTCTGAAG---AAGAAGTAA 2767

Db 848 GCATGGCACTACAGAGAAACAAACATTTAGAGATGTCCAGGAGAGCATGTAGTAA 907

Qy 2768 CATTTGACTTTTATCAAGGCTTCAAGATTTGGGAAGTTACGCCATGAAT-----A 2818

Db 908 CATTTACTCTTCACTCAGAGATATCAAAACCTGTGAAGCGCCAAAAAATTTCTCCTAGCA 967

Qy 2819 TACAAATGCTTTTGGAAAACTCAAGGAATTTCCCAAGTTAGAGAGCCAGAGACATGA 2878

Db 968 TACTAGCTATGCTGGAAACACTAGCAAAATGCTCCCTACCTAGAGTCCACAAAGACATGA 1027

Qy 2879 TAACGTGGATACTTCAGATGTTTCACACAGTGAAGCGATTAAAGAGAAAAATCTGTTTAA 2938

Db 1028 TTCGGTGGATTTGAAGACITTTTAATGCTGTAAAAAGATGAGGGAGA-----GTTTAC 1081

Qy 2939 TTGTAGCAACCATATCAGGATCGGAATCTATTAAAGATGATGAGATTACTCATGATAAAG 2998

Db 1082 CTACCAGTCCGTCGAGACACAGAGAACCATTAATGGAAGAGATTTCAAGGGACAAG 1141

Qy 2999 AAAAGCAGAACGAAAAAGAAAGCTGAAGCTGTAGCTACATCGCCAGAGATCATG 3058

Db 1142 ACAAGCTGAGAGGAAGAAAAAGCAGAGATTTGCCAGACTCGCCAGAGAAAAAGATCATG 1201

Qy 3059 CTCAGATGCTGCCCTTACAGAAAAAATTCATTGAAACTCAATAACTCATGTATGACATA 3118

Db 1202 CTCAGATGCTGTAATCGACGGCATTTTATTTGATGAAAAACAAAGAACTCTTTTCAGCAGA 1261

Qy 3119 CATCAGAAATGCTCGGAAAGAAAGATTCATTTATGGAGAGAGAGACCCACAGCAGTCA 3178

Db 1262 CATTAAGAACTGGATGCTCAACCTCTGCTGTGTTCTTGATCA-----TAGCCCTGTGGCTT 1315

Qy 3179 GTGACTACTCTAGAATTTGCTTTGGTCTCTAAACGGGGTCCATCTGTTACTGAAAAAGGAG 3238

Db 1316 CAGATATGACACTTACAGCACTTGGGCCCCGCAAACTCAGGTTCTGTGAACAAAGACAAT 1375

Qy 3239 TGCTCAGCTGCATCTTTGCCAAGAGAACAGGAGGTGAAATAGAAAAATATGCCATGG 3298

Db 1376 TCGTTACATGATATATGTTCAAGAGGACCAAGAGTTAAAGTGGAAACGAGGGCAATGG 1435

Qy 3299 TATTATCGCCCTGTGTCCAGAAATCTACTGCTTAAACCCAGCAGGGGAAAAACCCATAG 3358

Db 1436 TCTTGGCAGCATTTGTTTCAGAGATCAACTGTATTTATCAAAAAACAGAAAGTAAATTTATTC 1495

Qy 3359 RACTCTCAGGAGAACCCCTAGACCCACTTTTTCATGATCCAGACTTGGCATATGGAACIT 3418

Db 1496 AAGATCCAGAAAAATAT---GATCCATTAATTCATGCCCTCTGCTTGTGGAACAC 1552

Qy 3419 ATACAGGAAGCTGTGTCTATGTAATCCAGCAGTGTCTGGCAGAGATTTTGTGAAGCTG 3478

Db 1553 ACACCTAGTAGCTGTGGCAGCATTTGCAATGCCCATTTGTTGGCAAGGTATTTGATTCG 1612

Qy 3479 TACAGCTGAGCTCTCAGCAGCGCATTTGTTGACCTTTT-----TGACTTGG 3526

Db 1613 TTCAAGCTTAAAGAACAGGCAAGCAACAGAGATTTAGCTTTACATCAGAGCATGATGATAG 1672

Qy 3527 AAAGTGGAGATATCTTTCGCCCTCTTGCATATCTGTCATATCTGTCATCTGATCCCATTA 3586

Db 1673 AAAAGGAGATTTCTTTGCCCTTTTGTGAATGCTTGAATACTGTTTCTCTCTGCTG 1732

Qy 3587 TTCCCTTTCAGCTCAAAAGATAAACAGTGAAGATGCAGATGCTCTGCTCAACTTTTGA 3646

Db 1733 TGCCTT-----CTCCAAGAAATATTTTAAACACAGGTTAAATTTTTCAGACCAACCAA 1786

Qy 3647 CCTCGCAGGTTGGATACAGACTGTCTGGCCAGAAATATCAGGTTTATAATATAAGACATG 3706

Db 1787 ATCTGACTCAGTGGATTTAGAACAAATATCTCAGCAATAAAGCATTTACAGTTTCTTAGGA 1846

Qy 3707 CTAAAGGAGAAACCCAAATTCCTATTTCTTTAAATCAGGAATGGAGATTTCTACTTTGG 3766

Db 1847 AAGAAAGAAAGTACTCTTAATAATGCTCTACAAAGAAATTCAGAAAAATGGGATGAATAC 1906

Qy 3767 AGTTCCATTCCTCAGTGTGCTGAGTTTGGGCTTTCAGTCTTCGATTAATAATTTCAAAATAGCATCA 3826

Db 1907 AGCTCCCTGAAGGGTTTCAGGCTGATTTTCGCTCTAAGATCCCTTATTTCTGAGAGCATTA 1966

Qy 3827 AGGAAATGTTTATCTCTTTGCCACAAATTTTATAGAATTTGATTTGAAGTGGCCACTG 3886

Db 1967 AAGAAATGCTAACGACATTTGGAACCTGCTACCTACAAAGTGGGACTTAAAGGTTTCATCCCA 2026

Qy 3887 ATGAAAGGATCCTCGAGTCCCATCTGACCTGGAGACCTGCGCTTTTCACTATCATCAGG 3946

Db 2027 ATGAGAGAGATCTCGTGTTCCTCAATATGTTGGGTAGCTGCGGCTACACCATCAAA 2086

Qy 3947 CAATTTGAAATCTATTGGAGATGAAGAAAAACCTCTGTTTGGAGCATTCAAAATAGGC 4006

Db 2087 GCATAGAAAGAAATTTGATGATGAAGATAACCATTTGTTGGTCTTTTACCTTTGAGAC 2146

Qy 4007 AGCATTAATGGTCTGAAAGCATTAATGCACTTTTCAGTTTGCACAGAGATTTACCTGCTCTC 4066

Db 2147 TGGATGACTGTCTTAGTTCATTGACGAGATTTGGCCGACGACACACTGGACAGTGGCATCAG 2206

Qy 4067 AGGTCCTGATACAGAAAAATCTGGTTCGCTCTATCATGTTGTTCTTCTCTTAACATAAAT 4126

Db 2207 TTTTCAGTGGTGAAGACATTTTGTAAACTTTTTCATCACTGCTGCTGCTTAATGAGACGC 2266

Qy 4127 CAGAGATACACCATGCTCTGCTATAGATCTGTTTTCATGTTTGGTGGGCTGCTGTGT 4186

Db 2267 ATGAGGAACCTTCCATGCATATTAGATATTGACATGTTTATTTATTTGGTGGGCTTGGTGC 2326



Db	568	AGATACAGACATCCCACTCCCGGTGGTGGCTCCATTCCTGCGCCTCTGTGTTGCAAGCGCTGG	727
QY	2591	TTAACTCTTCAACTGTGATATCATGATGTACATCTCTCAGAGCCGCTATTGTAGCGGGCAA	2650
Db	728	TTRACATTTTGCAGTCAGATGTCATGTTGTGTCATCATGGGAACAATCTCCATGGGCTG	787
QY	2651	TAGACAGATTTCTAATCTTGTGGACCGAAGGATGCTCCAAATGGCTTTTCATATCTCTGG	2710
Db	788	TGGAACATAATGGATATGCTGCTGTCAGATGCCATGCTGCAAGGGTGTACATTTAATTG	847
QY	2711	CATTGGGTTTACTAGAGAGAACACAGCTTCCAAAAGCTCTCTGAAG---	2767
Db	848	GCATGGACATACAGAAGAAAACAATTTAGAGATGTCACGGAAGCATGTAGTAA	907
QY	2768	CATTGTACTTTTATCATAGGCTTCAAGTTGGAAGTTCCAGCCATGAAT-----A	2818
Db	908	CATTACCTTCACTCAGAGATATCAAAACCTGGTGAAGGCCCAAAAATTCCTTAGCA	967
QY	2819	TACAAATGCTTTTGGAAAACACTCAAAGGAATTTCCCGAGTTTCCAGGCCAGAGACATGA	2878
Db	968	TACTAGCTATGCTGGAAACACTACAAAATGCTCCCTACCTTGAAGTCCCAAGACATGA	1027
QY	2879	TAAAGTGGATCTCAGATGTTTGCACACAGTGAAGCGATTAAGAGAAAATCTGTGTTAA	2938
Db	1028	TTCCGTTGGATTTTGAAGACTTTTAAATGCTGTTTAAAAGATGAGGGAGA-----GTTCA	1081
QY	2939	TTGTAGCAACCATCAGGATCGGAATCTATTAAAGATGATGAGATTACTCATGATAAAG	2998
Db	1082	CTACCACTCCGTGGCAGACAGAGAGGAACCATATGGAAGAGGTTCAAGGGACAAG	1141
QY	2999	AAAAAGCAGAACGAAAAGAAAAGCTGAAGCTGCTAGGCTTACATCGCCAGAGACATGG	3058
Db	1142	ACAAAGCTGAGGAAGAAAAGCAGAGATTCAGAGCTGCGCAGAGAAAAGATCATGG	1201
QY	3059	CTCAGATGCTGCTTTACAGAAAACCTTCATGAAACTCATAACTCATGTATGACAATA	3118
Db	1202	CTCAGATGCTGAAATGCGAGCGCATTTTATTGATGAAAACAAGAACTCTTTTCAGCAG	1261
QY	3119	CATCAGAAATGCTGGGAAGAAGATTCCATTTATGAGGAAGAGACACCCAGCAGTCA	3178
Db	1262	CATTAGACTGGATGGCTCAACCTGCTGCTGTTCTTGATCA-----TAGCCCTGGCTT	1315
QY	3179	GTGACTACTAGAAATTTGGTGGTCCCTAAACGGGGTCCATCTGTTACTGAAAGGAGG	3238
Db	1316	CAGATATGACACTTACAGCACTGGCCGCCACAACTCAGGTTCTGAAACAAGACAAT	1375
QY	3239	TGCTGAGTGCATCTTTGGCCAGAGAAGACAGAGGTGAAATAGAAAATATGCCATGG	3298
Db	1376	TCGTTACATGATATTGTGTCAGAGAGGCAAGAAGTTTAAAGTGGAAAGCAGGGCAATGG	1435
QY	3299	TATTATCGGCTGTGTCAGAAATCTACTGCTTTAACCCAGCACAGGGGAAAACCCATAG	3358
Db	1436	TCCTGGCAGATTTGTCAGAGATCAACTGATATTATCAAAAACAGAAATTTATTTC	1495
QY	3359	AACTCTCAGGAGACCCCTAGACCCACTTTTCATGATCCAGACTTGGCATATGGAACCT	3418
Db	1496	AAGATCCAGAAAATAAT---GATCCATATTATCATGACCCCTGATGCTCTGTGGGAAC	1552
QY	3419	ATACAGAACTGTGGTCATGTAATGACCGCACTGTCTGGCAGAGATATTTTGAAGCTG	3478
Db	1553	ACACTAGTACTGTGGGCACATTTATGATGCCCATTTGTTGGCAAGGATTTTGTATTCG	1612
QY	3479	TACAGCTGAGCTCTCAGCAGCGCATTCATGTTGACCTTTT-----TGACTTGG	3526
Db	1613	TTCAAGCTAAAGAACAGCGAAGGCACAGAGATTACGCTTACATACGAGCTATGATGTAG	1672
QY	3527	AAAGTGAGAAATATCTTTGGCCCTCTTGGAAAATCTCTGTGCAATACTGTGATCCCATTA	3586
Db	1673	AAAACGGAGATTTCTTTGGCCCTTTGTGAATGCTTGAGTAATACTGTTATTCTCTG	1732
QY	3587	TTCCCTTTGCAACCTCAAAGATAAACAGTGAGAAATGCAGATGCTTTGCTTCAACTTTGA	3646
Db	1733	TGCTTT-----CTCCAAGAAATATTTTAAACACAGGTTAAATTTTTCAGACCAACCAA	1786

Qy	4724	AAAGAAATAGTTTGATAGAGCTTCCCTGATGACTATAGTCGCTCCTGAATCAAGCTTCTC	4783
Db	2849	AATCTAACAATAATTAACCTTCCAGAGATTACAGCAGCTCATTAATCAAGCATCCA	2908
Qy	4784	ATTTFCAGGTGCCACGGTCTGCAGATCATGAGCGAAAGCATCTGTCTCCTGCTGCTTTCT	4843
Db	2909	ATTTCCTGTCGCCAATACAGTGGTGATTAAGACAGACGCCCACTCTGTGCTTGTGT	2968
Qy	4844	GTTGGGCTATFACATATGTTTCAGAACAAITTTGCTGCCAGGAAATTTGTGAACGGGGAAGAGG	4903
Db	2969	GCGSATCTCGCTGTGCTCCAGAGTTTACTTGCTGCCAGACTGAACCTGGAAGGGAGGATG	3028
Qy	4904	TTGSAGCTTCGCATTTTTCAGSCACTTCACCTGTSGAGCCGGAGTCTGCATTTTCTTAAAA	4963
Db	3029	TAGGAGCCTGCACAGCTCACACCTACTCTCTGTGGCTCTGGAGTGGGCACTTCTCTGAGAG	3088
Qy	4964	TCAGAGAATGCCAGTGGTCTCTGGTTGAAGGTAAAGCCAGAGGCTGTGCTATCCAGCTC	5023
Db	3089	TACGGGAATGTCAGGTGCTATTTTTTAGCTGCGAAAACCAAAGGCTGTTTTATTCCTCTC	3148
Qy	5024	CTTACTTTGGATGAATATGGAGAACACACCCTGGCTGTGAAGAGGGGCAACCCCCTTCATT	5083
Db	3149	CTTACCTTGATGACTATTTGGGAGACCGACAGGAGACTCAGACGGGGAAATCCTTTACATT	3208
Qy	5084	TATCTCTGAGCGGTATCGGAAGCTCCATTGTTGTTCTGGCAACAACACTGCATTATAGAAG	5143
Db	3209	TATGCAAAAGACGATTCAGAGAATTCAGAAAGCTCTGGCAACACACAGTGTCCACAGG	3268
Qy	5144	AGATTGCTAGGAGCCCAAGAGACTAATCAGATGTTATTTTGGATTCAACTGGCA	5195
Db	3269	AAATTGGACATGCACAGGAAGCAATCAGACACTGGTTGGCACTTGACTTGATGGCA	3320

### RESULT 3

```

US-09-529-063-57
; Sequence 57, Application US/09529063
; Patent No. US20020102542A1
; GENERAL INFORMATION:
; APPLICANT: FUKUSHIMA, DAIKICHI
; APPLICANT: SHIBAYAMA, SHIRO
; APPLICANT: TADA, HIDEAKI
; TITLE OF INVENTION: POLYPEPTIDE, CDNA ENCODING THE POLYPEPTIDE, AND USE OF
; TITLE OF INVENTION: THE BOTH
; FILE REFERENCE: Q58769
; CURRENT APPLICATION NUMBER: US/09/529,063
; CURRENT FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: PCT/JP98/04514
; PRIOR FILING DATE: 1998-10-06
; PRIOR APPLICATION NUMBER: JP 9-274674
; PRIOR FILING DATE: 1997-10-07
; NUMBER OF SEQ ID NOS: 117
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 57
; LENGTH: 3502
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: (56)..(109)
; NAME/KEY: mat_peptide
; LOCATION: (110)..(3382)
; NAME/KEY: CDS
; LOCATION: (56)..(3382)
US-09-529-063-57

```

QY	1952	GAATGGAGCTGCTCTCTATTAGCCAGGCTGTTTTATTACCAAGATGTTAAAGTCGAGAAG	2011
Db	123	GAATGGGTTCTCTCTAGTAACACGATTTATTACTACCATAAATGTGAATTCGACAGCTG	182
QY	2012	AAATGTATGATAAGAGATATCATCTTTAGATTTGGTGCATCTTTAATGGATCCCAATA	2071
Db	183	AGATGTTTGACAAGGATGTAGTAATGCTTCAGACAGGTGCTCCATCATGGATCCAAATC	242
QY	2072	AGTTCTTTACTGGTACTTCAGAGGTATGAACCTTGCAGGCGCTTTAAACAAGACCATAT	2131
Db	243	ATTTCCTGATGATCATGCTCAGCGCTTTGAACCTTTATCAGATTTTCAGTACTCCAGACT	302
QY	2132	CTACAAA-----GACCAGGATTTGATTAACAATATAATATAC	2170
Db	303	ATGAAAAGAGATTTAGTTCTCAGATTAACCATAGGATGTTGTTTCAGCAAGAACAAATACT	362
QY	2171	TAATAGAGAAATGCTTCAGGCTCTCATCTATATTGTTGGTGAGCGTTATGTACCTGGAG	2230
Db	363	TAATAGAGAAATGCTTACCTCATTAATATGCTTTGGAGAGAGATTTAGTCCCTGGAG	422
QY	2231	TGGAAATGTGACCAAAAGAGAGTGCACAAATGAGAGAAATCATTCACCTTGCTTGCAATTG	2290
Db	423	TTGACAGGTAATGCTTACAGATGAATCAAGCGAGAGATTATCCATCAGTTGAGTATCA	482
QY	2291	AACCCATGCCACACAGTGCCCATGCCAAAATTTTACCTGAGAAATGAAAAATTAATGAACCTG	2350
Db	483	AGCCTATGGCTCATAGTGAATGGTAAGTCTTTTACCTGAAGATGAGAACAAAGGAGACTG	542
QY	2351	GCTTAGAGAAATGCTATAAACAAAGTGCCACATTTAAGAACACAGGTGTATCAGGCCATG	2410
Db	543	GCATGGAGAGTGAATCGAAGAGCTGCCCATTTCAAGAAACCTTGGATTAACAGGACGAG	602
QY	2411	GAGTTTATGAACATAAAGATGAATCACTGAAAGACTTCAATATGCTACTTTTATCATTA	2470
Db	603	GCATGTATGAATGAACACCAAGATGTGCCAAGAGATTCACTTGATTTCTATCACATTT	662
QY	2471	CCAAAACCCAGCATAGCAAGGCTGAACATATGCAAGAAAGAGAGAAAAACAAGAAAAA	2530
Db	663	CAAGGCGAAGACAGTCCAAAGCAGAAAGCGCAACGGAATTTGAAAAGACAAAAATAGAG	722
QY	2531	AAGATGAAGCATTCGCCCCACCACCTCCTGAATTTCTGCCCTGCTTTCAGCAAGATGA	2590
Db	723	AAGATACAGCACTCCCACTCCGGTGGCTCCTCATTTGCGCTCTGTTTGCAGCGCTGG	782
QY	2591	TTAACTTCTCAACTGTGATATCATGTATCATTTCTCAGGACCGTATTTGAGCGGCA	2650
Db	783	TTAACTTTGCAGTCAGATGCTATGTTGTCATCATGGGAACAATTTGCAATGGGCTG	842
QY	2651	TAGACAGATTTAACTTTGGACCGAAGGATGCTCCAAATGGCTTTTCATATCTCGG	2710
Db	843	TGGAACATAATGGATATGCTCTGGTCAGAGTCCATGCTGCAAGGGTGTACATTTAATTG	902
QY	2711	CATTGGGTTTACTAGAGAGAAAGCAACAGTTTCAAAAAGCTCTCTGAAG---AAGAATAA	2767
Db	903	GCATGGCACTACAAGAGAAACCAACATTTAGAGAAATGTACCGAAGACGATGTAGTAA	962
QY	2768	CATTGTACTTTTATCATAAAGCTTCAAGATTGGGAAGTTCAGCCATGAAT-----A	2818
Db	963	CATTTCCTCACTCAGAGATATCAAAACCTGTTGAAGCCCAAAAATTTCTCTAGCA	1022
QY	2819	TACAAATGCTTTTGGAAAACCTCAAGGAATTTCCCCAGTTTAGAAGGCCAGAGACATGA	2878
Db	1023	TACTAGCTATGCTGSAACACTACAAAATGCTCCCTACCTAGAAAGTCCCAAGACATGA	1082
QY	2879	TAACTGGATPACTTCAGATGTTTGACACAGTGAAGCGATTAAAGAGAAAAATCTTTTAA	2938
Db	1083	TTCCGTGGATATTGAAGACTTTTAAATGCTGTAAAAAGATGAGGAGA-----GTTTCA	1136
QY	2939	TTGTAGCAACCATCATAGGATCGGAATCTATTAAAGATGATGAGATTACTCATGATAAG	2998
Db	1137	CTACAGTCCCGTGGCAGACAGAGGAACCATTAATGGAAGAGAGTTCAAGGGACAAG	1196
QY	2999	AAAAAGCAGAACAAAAAGAAAGCTGAAGCTGTAGCTACATCGCCAGGAAGATCATGG	3058

1197 ACAAGCTCAGAGGAGAGAAAGCAGAGATTGCCAGACTGCCGAGAGAAAGATCATGG 1256  
3059 CTCAGATGCTGCCTTTACAGAAAACCTTCATTGAAACTCATAACTCATGTATGACAATA 3118  
1257 CTCAGATGCTGAAATGCAGCGCATTTTATTGATGAAAAACAAGAACTCTTTTCAGCAGA 1316  
3119 CATCAGAAATGCTCGGAAAGAAATCCATTATATGAGGAGNAGAGACCCCGAGCAGTCA 3178  
1317 CAATTAGAACTGGATGCTCAACCTCGCTGCTTTCATCA-----TAGCCCTGTGGCTT 1370  
3179 GTGACTACTCTAGAAATGCTTTGGGTCCTTAAACGGGGTCCATCTGTACTGAAAGAGG 3238  
1371 CAGATATGACACTTACAGCACTGGGCCCGCAGCAAACTCAGGTTCTGTACAAAGACAAT 1430  
3239 TCGTGACGTGCTCCTTTGCCAAGAAAGACAGGAGGTGAAATAGAAAATTAATGCCATGG 3298  
1431 TCGTTACATGTATATGTCTCAAGAGGAGCAAGAAGTTAAAGTGGAAAGCAGGGCAATGG 1490  
3299 TATTATCGGCCTGTGCCAGAAATCTACTGCTTAACCCAGCAGCAGGGGAAACCCATAG 3358  
1491 TCTTGGCAGCATTTGTTACAGATCAACTGTATTATCAAAAACAGAAAGTAAATTTATTC 1550  
3359 AACTCTCAGGAGAACCCCTAGACCCACTTTTCATGGATCCAGACTTGGCATATGGAACCT 3418  
1551 AAGATCCAGAAAATAT---GATCCATTAATTCATGCACCCCTGATCTGTGTGGAAAC 1607  
3419 ATACAGGAAGCTGTGGTCAATGTAATGCACGCACTGTGCTGGCAGAAAGTATTTGAAGCTG 3478  
1608 ACACTAGTAGCTGTGGGCACATATATGCATGCCCAATTGTTGGCAAGGTATTTGATTCG 1667  
3479 TACAGCTGAGCTCTCAGCAGCGGATTCATGTTGACCTTTT-----TGACTGG 3526  
1668 TTCAAGCTAAAGAACAGCAGGAAGCAACAGAGATTACGCTTACATACGAGCATATGATAG 1727  
3527 AAAGTGGAGAAATCTTTGCCCTCTTTGCCCTCTGTGCAATCTGTGCAATCTGTACCCCATTA 3586  
1728 AAACGGAGAAATCTTTGCCCTCTTTGGAATGCTTGAGTATACCTGTTATTCCTCTGC 1787  
3587 TTCCCTTTGCACTCAAAAGATAAAGAGTGAAGTGAAGTGCATGCTCTGTGCTCAACTTTGA 3646  
1788 TGCTTT-----CTCCAAGAAATATTTTAAACAACAGGTTAAATTTTTCAGACCAACCA 1841  
3647 CCCTGGCAGGTGGATACAGACTGTCTGGCCAGAAATATCAGGTTATATATAAGACATG 3706  
1842 ATCTGACTCAGTGGATTAGAACAAATATCTCAGCAAAATAAAGCAATACAGTTTCTTAGGA 1901  
3707 CTAAGGAGAAACCCAAATCTCTATTTCTTTTAACTCAAGGAATGGGAGATTCTACTTTGG 3766  
1902 AAGAGAAAGTACTCTTAATATGCTCTACAAAGAAATTCAGAAATGTGGATGAATAC 1961  
3767 AGTTCCATTTCCATCTGAGTTTGGGTTGAGTCTTTCGATTAAATATTCAAATPAGCATCA 3826  
1962 AGCTCCCTGAAGGTTTCAGGCTGATTTTCGCTTAAAGTCCCTTATCTGAGAGCATAA 2021  
3827 AGAAATGTTATCTCTTTGGCACAACAATTTATAGAAATGGAATGGAAGTGCACCTG 3886  
2022 AAGAAATGCTAAAGCATTTGGAATGCTACCTACAAAGTGGGACTTAAAGTTTCATCCCA 2081  
3887 ATGAAGGATCTCGAGTCCCATGCTGACCTGGAGCAGCTGCGCTTTTCACTATCCAGG 3946  
2082 ATGAGAGAGATCTCGTGTTCCTCAATGTTGGGTTAGCTGCGGTTACACCATCCAA 2141  
3947 CAATTGAAATCTATTGGGAGATGAAGAAACCTCTGTTTGGAGCATTCAAAATAGGC 4006  
2142 GCATAGAAAGAAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2201  
4007 AGCATAATGGTCTGAAGCAATTAATGCAATTTGCACTTGCACAGAGGATTAATCTGCTC 4066  
2202 TGGATGACTGTCTTAGTCTATTGACGAGATTTTCCCGCAGCAGCACTGGACAGTGGCATCAG 2261  
4067 AGGTCTGTATCAGAAACATCTGGTCTCTCTATCTATCAGTTGTTCTCTCTCAACATAAAT 4126

2262 TTTCAGTGGTGCAAGGACATTTTTTGTAAACTTTTTTGCATCACTGGTGCCTAATGACAGCC 2321  
4127 CAGAAGATACACCATGCTTCTGTCTATAGATCTGTTTTCATGTTTGGTGGTGTGTGT 4186  
2322 ATGAGAACTTCCATGATATATAGATATGACATGTTTCAATTTATTTGGTGGCTTGGTGC 2381  
4187 TAGCATTTCCCATCTTCTGTTATTTGGGATGACCTGTTGATCTCTCAGCCCTTCTTCAGTTAGTT 4246  
2382 TTGCATTTCTGCTGCTTCAGTG-----TCAGGATTTTTCAGGGATCAGCC 2426  
4247 CTTCTCTAATACCACTTTTATCTCTTCCATTTGATCACCATGGCAGACATGTTCCAGATAC 4306  
2427 TTGGCACTGGAGACCTTCACATTTTCCATCTGTTTACTATGSCACACATCATACAGATCT 2486  
4307 TACTTACAGTAGACAC---AGGCCATCCCTTGTCTAGGTTTCAAGAAGACAGTGAAGAGG 4363  
2487 TACTTACCTCATGTACAGAAGAGAAATGSCATGGATCAAGAAATATCCCTTTGTGAAGAAG 2546  
4364 CTCAATCCGCACTCTCTTTTCTTTCAGAAATTTCTCAATATATACAGTGGCTTCAATGGGT 4423  
2547 AATCAGCAGGTTCTTGTCTTTGTATATAACACTTCCACAGTATACGGGAAGTGCCTTGAAG 2606  
4424 GTGATATTCCTGGCTGCTATTTTGGGCTCTCACTGAAGATGGCATCACCCCTTATCTTC 4483  
2607 AATACCATCCGGCTGCATCTGTGGAGAGTGTACAGCTGGAATCATGCTTTCCTGA 2666  
4484 GCTGTGCTGCTATTTTTCACATTTTCTTGGGTAACCTCCGCTGAGAACTGCATA 4543  
2667 AGTGTCTGCTTTATTTTTCATTTAAATGAGTTTCTTCCCCACCCGACATTCAG 2726  
4544 CCAATTCGAGAAAGAGAGTACAGTGCATCTCTAGTATCTATCTTACTTACTACATAAT 4603  
2727 T---TCCTGGAAACAAGCCATTTTGAACATTTATGTAGTATCTTTCCTACCAACAAC 2783  
4604 TGTTCCTGCTCTCCAGGAATATTTGGGATCTGTAAGCCCTCTGCTCAGAGCGGTGTG 4663  
2784 TCATTTGCCCTTTTCAAGAAATAGTGAGATAATGAAATCTACTGATTGAAAGTTGGTGCC 2843  
4664 CAGATCCTGCTTACTAACTGTTTGAAGCAAAAACACCGTGTGCTCAGGTACCTAGAA 4723  
2844 GTAACAGTGAAGTTAAAGATATCTAGAAGGTGAAGAGATGCTATAGATATCCAAGAG 2903  
4724 AAGAATAAGTTTGAAGAGCTTCTCTGATGACTATAGTGTCTCTGATCAAGTCTCTC 4783  
2904 AATCTAACAATTAATAAACCTTCCAGAGGATTACAGCAGCTCATTAATCAAGCATCCA 2963  
4784 ATTTCAAGTCCCAAGCTGTCAGATGATGAGCAAGACATCTCTCTGCTCTTCTTCT 4843  
2964 ATTTCTGCTGCCGAATCAGGTGTTGATAGACAGAGCCCAACTCTGTGCTTGTGT 3023  
4844 GTGGGCTATCTATGTTCTCAGAAACATTTTCTGCCAGGAAATTTGTGAACGGGAAGAGG 4903  
3024 GCGATCTCTGCTGTCTCCAGAGTTTACTGCTCCAGACTGAACCTGAAGGGGAGGATG 3083  
4904 TTGAGCTTGCATTTTTCAGCACTTCTGCTGAGCGCGGAGTCTGATCTTCTTCTTCTTCT 4963  
3084 TAGGAGCTGCACAGCTCAGACCTTCTCTGCTGCTGAGTGGAGTGGCATCTTCTCTGAGAG 3143  
4964 TCAGAGAAATCCGAGTGGTCTCTGTTGAAGTGAAGCAAGAGGCTGTGCTTATCCAGCTC 5023  
3144 TAGCGGAATGTCAAGTCTATTTTGTGCAAAACCAAGGCTGTTTTTCTCTCTC 3203  
5024 CTTTACTTGGATGATAGGAAACAGACCTGCTGCTGAAGGGGCAACCCCTTCTTCT 5083  
3204 CTTTACTTGTATGATAGTGGGAGACCCAGGAGCTCAGAGGGGAATCTTTCATTT 3263  
5084 TATCTCTGAGCGGTATCGGAAGCTCCATTTGCTGCGCAACAACATGCAATTTATAGAG 5143  
3264 TATGCAAGAGAGGATTCAGAAAGATTCAGAGCTCTGCGACCAACACAGTGTACAGAGG 3323  
5144 AGATTGCTAGGAGCAAGAGACTAATCAGATGTTTATTTGGATTTCAACTGGCA 5195  
3324 AATTTGACATGCACAGGAAGCCATCAGACACTGGTTGGCATTTGACATGGCA 3375





Qy	3419	ATACGAAAGCACTGTGGTCATGTAATGCACGCACTGTGCTGGCGAAGATATATTTTGAAGCTG	3476
Db	1608	ACACTAGTACTGTGGGCACATTAATGCATGCCCATGTTGGCAAGGATATTTTGATTCGG	1667
Qy	3479	TACAGCTGAGCTCTCAGCAGCGCATTCATGTTGACCTTTT-----TGACTTGG	3526
Db	1668	TTCAAGCTAAGAACACGAGNAAGGCAACAGAGATTACGCTTACATACGAGCTATGATGTAG	1727
Qy	3527	AAAGTGGAGAAATATCTTTGGCCCTCTTGTGAAATCTCTGTGCAATACTGTGATCCCATTA	3586
Db	1728	AAAACGGAGAAATTCCTTTGGCCCTTTGTGAATGCTTGAGTAACTGTTATTTCCCTCTCG	1787
Qy	3587	TTCTTTTGGCAACTCAAAAGATAAACAGTGAGAAATGCAGATGCTTTGCTCAACTTTTGA	3646
Db	1788	TGCTTT-----CTCCAAGAAATATTTTTAAACAACAGGTTAAATTTTTCAGACCAACCA	1841
Qy	3647	CCCTGGCACGGTGGATACAGACTGTTCTGGCCAGAAATATCAGGTTATATATTAAGACATG	3706
Db	1842	ATCTGACTCAGTGGATTAGAACAAATATCTCAGCAAAATAAAGCATTTACAGTTTCTTAGGA	1901
Qy	3707	CTAAAGGAGAAAACCCAATTCCTATTTCTTTAATCAAGGAATGGGAGATTCTACTTTGG	3766
Db	1902	AGNAGAAAGTACTCTCTAATAATGCCTCTACAAGAAATTCAGAAAATGTGGATGAATTAC	1961
Qy	3767	AGTTTCCATTCCATCCTGAGTTTTGGCGTTTGAGTCTTCGATTAATATATCAAAATAGCATCA	3826
Db	1962	AGCTCCCTGAAGGGTTCAGGCCCTGATTTTCGTCTTAAGATCCCTTATTTCTGAGAGCATAA	2021
Qy	3827	AGGNAATGGTTATTCCTTTGGCCACAACAATTTATAGAATTTGGATTGAAAGTGCCACCTG	3886
Db	2022	ARGAAATGCTTAACGACATTTTGGAACTGCTACCTTACAAGGTGGGACTTAAAGGTTTCATCCCA	2081
Qy	3887	ATGAAAGGATCCTCGAGTCCCATCTGACCTGGAGCACCCTCGCTTTTCACTATATCCAGG	3946
Db	2082	ATGNAGAGGATCCTCGTGTTCCTCAATAATGTGTGGGGTACCTGCGCGTACACCATCCAAA	2141
Qy	3947	CAATTGAAAATCTATTGGGAGATGAAGAAAACCTCTGTTTGGAGCACTTCAAAATAGGC	4006
Db	2142	GCATAGAAAGAAATTTTGAGTGATGAAGATAAACCATTTGTTGGTCTCTTACCTTGCAGAC	2201
Qy	4007	AGCATATGCTCGAAGCAATTAATGCAGTTTTCGAGTTGCACAGAGAGATTACCTGTCCCTC	4066
Db	2202	TGGATGACTGCTTAGTGCAATTACAGAGATTTGCCGCAAGCACTGGACAGTGGCATCAG	2261
Qy	4067	AGGTCCTGATACAGAAACATCTGGTCTCTATCATAGTTGTCTTCTCTTAACATAAAAT	4126
Db	2262	TTTCAGTGGTGCAGGACATTTTGTAAACTTTTTGCATCACTGGTGCCCTAATGACAGCC	2321
Qy	4127	CAGAAGATACACCATGCCCTTCTCTATAGATCTGTTTTCATGTTTGGTGGGTGCTGTGT	4186
Db	2322	ATGAGGAACCTCCATGCATATTAGATATTGACATGTTTCATTTATTGGTGGGCTTGGTGC	2381
Qy	4187	TAGCATTTCCATCCTTGTAATTGGGATGACCCCTGTTGATCTGACGCTCTTCTAGTTAGTT	4246
Db	2382	TTGCATTTCTCGGTTGCACTG-----TCAGGATTTTTCAGGGATCAGCC	2426
Qy	4247	CTTCTATACCCACCTTTATCTCTTCCATTTGATACCATGGCACACATGCTTCACATAC	4306
Db	2427	TGCGCACTGGAGACCTTCACATTTCCATCTGGTTACTTATGGCACACATCATACAGATCT	2486
Qy	4307	TACTTTACAGTAGACAC---AGGCCTACCCCTTGCTCAGGTTTCAAGAAGACAGTGAAGAGG	4363
Db	2487	TACTTACTCTATGTACAGAGAGANYGCGATGGATCAAGAAAATCCCCCTTGTGGAAGAAG	2546
Qy	4364	CTCATTCGGCATCTCTCTTTCTTTCGAGAAATTTCTCAATATACAAGTGGCTCCATTTGGT	4423
Db	2547	AATCAGCAGTTCTTGCTTTGTATAAAACACTTCACCAGTATACGGGAAGTGCCCTTGAAG	2606
Qy	4424	GTGATATTCCTGGCTGATTTTGGGGTCTCACTGAAGAAATGGCATCACCCCTTATCTTC	4483
Db	2607	AAATPACCATCCGGCTGCACTCTGTGGAGGAGTGTCAAGAGCTGGAATCATGCTCTTCTGA	2666
Qy	4484	GCTGTGCTGATGTTTTTTTCCACTATTATTACTTTGGGGTAACTCCGCTGGAGAACTGCATA	4543

RESULT 5

RESULT 5  
US-09-918-995-36721  
; Sequence 36721, Application US/09918995  
; Publication No. US20030073623A1

```

? GENERAL INFORMATION:
?
? APPLICANT: Hyseq, Inc.
?
? TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
?
? TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
?
? FILE REFERENCE: 204111-756
?
? CURRENT APPLICATION NUMBER: US/09/918,995
?
? CURRENT FILING DATE: 2001-07-30
?
? PRIOR APPLICATION NUMBER: US/09/235,076
?
? PRIOR FILING DATE: 1999-01-20
?
? NUMBER OF SEQ ID NOS: 38054
?
? SOFTWARE: FastSeq for Windows Version 3.0
?
? SEQ ID NO 36721
?
? LENGTH: 505
?
? TYPE: DNA
?
? ORGANISM: Homo sapiens
?
? FEATURE:
?
? NAME/KEY: misc_feature
?
? LOCATION: (1)...(505)
?
? OTHER INFORMATION: n = A,T,C or G
?
? US-09-918-995-36721

```

Query Match 8.4%; Score 438.4; DB 11; Length 505;  
Best Local Similarity 99.8%; Pred. No. 2.7e-110;  
Matches 439; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3834 GGTATTCTCTTGGCCACAAATTTATAGAAATTTGGAATGGAAGTGCACCTGATGAAG 3893  
|||||  
Db 65 GGTATTCTCTTGGCCACAAATTTATAGAAATTTGGAATGGAAGTGCACCTGATGAAG 124  
|||||

QY 3894 GGATCCTCGAGTCCCCATGCTGACCTGGAGCACCTGGCTTTCACTATCCAGGCAATTGA 3953  
|||||  
Db 125 GGATCCTCGAGTCCCCATGCTGACCTGGAGCACCTGGCTTTCACTATCCAGGCAATTGA 184  
|||||

QY 3954 AAATCTATTGGGAGATGAAGGAAAAACCTCTGTTTGGAGCACTTCAAAATAGGCGAGCATAA 4013  
|||||  
Db 185 AAATCTATTGGGAGATGAAGGAAAAACCTCTGTTTGGAGCACTTCAAAATAGGCGAGCATAA 244  
|||||

QY 4014 TGCTCTGAAGACATTAATGAGTTTGCAGTTTGCAGATGTCACAGAGGATTACCTGTCCTCAGGTCT 4073  
|||||  
Db 245 TGCTCTGAAGACATTAATGAGTTTGCAGTTTGCAGATGTCACAGAGGATTACCTGTCCTCAGGTCT 304  
|||||

QY 4074 GATACAGAAACATCTGTTTGGCTTCTCTATCAGTTTGTCTTCCCTAACATAAAATCAGAAAGA 4133  
|||||  
Db 305 GATACAGAAACATCTGTTTGGCTTCTCTATCAGTTTGTCTTCCCTAACATAAAATCAGAAAGA 364  
|||||

QY 4134 TACACATGCTCTGCTCTATAGATCTGTTTCATGTTTGGTGGGTGCTGTGTTAGCAT 4193  
|||||  
Db 365 TACACATGCTCTGCTCTATAGATCTGTTTCATGTTTGGTGGGTGCTGTGTTAGCAT 424  
|||||

QY 4194 CCATCCTTGTATTGGGATGACCTCTGTGATCTGCAGCTTCTTCAGTTAGTTCTTCTCTA 4253  
|||||  
Db 425 CCATCCTTGTATTGGGATGACCTCTGTGATCTGCAGCTTCTTCAGTTAGTTCTTCTCTA 484  
|||||

QY 4254 TAACACCTTTTATCTCTTCC 4273  
|||||  
Db 485 TAACACCTTTTATCTCTTAC 504  
|||||

## RESULT 6

US-09-822-849A-532  
; Sequence 532, Application US/09822849A  
; Patent No. US20020045170A1  
; GENERAL INFORMATION:  
; APPLICANT: Wong, Gordon G.  
; APPLICANT: Clark, Hilary  
; APPLICANT: Fechtel, Kim  
; APPLICANT: Agostino, Michael J.  
; APPLICANT: Howes, Steven H.  
; APPLICANT: Resnick, Richard J.  
; APPLICANT: Gulukota, Kamalakara  
; APPLICANT: Graham, James R.  
; APPLICANT: Genetics Institute, Inc.  
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS  
; FILE REFERENCE: GIN 6403  
; CURRENT APPLICATION NUMBER: US/09/822,849A  
; PRIOR FILING DATE: 2001-09-04  
; PRIOR APPLICATION NUMBER: 60/195,582  
; PRIOR FILING DATE: 2000-04-06  
; NUMBER OF SEQ ID NOS: 598  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 532  
; LENGTH: 972  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-822-849A-532

Query Match 4.8%; Score 249; DB 9; Length 972;  
Best Local Similarity 56.8%; Pred. No. 1.1e-57;  
Matches 478; Conservative 0; Mismatches 360; Indels 3; Gaps 1;

QY 4355 GTGAAGAGGCTCATTCGGCATCTCTTCTTTGTCAGAAATTTCTCAATATACAGTGGCT 4414  
|||||  
Db 15 GTGAAGAGGAATCAGCAGTCTTCTGCTTTGTATATAAACACTTCACCATATACGGGAAGTG 74  
|||||

QY 4415 CCATTGGGTGTGATATTCTCTGGTGTGTTTGTGGTCTCACTGAAGAAATGGCATCACCC 4474  
|||||

Db 75 CCTTGAAGAAATACCATCCGCTGGCATCTGTGGAGGAGTGTACAGAGCTGGAAATCATGC 134  
|||||

QY 4475 CTTATCTTGGCTGTGCTGCATTTTTCCTACTATTTTACTTGGGGTAACCTCCGCCCTGAGG 4534  
|||||

Db 135 CTTTCTCTGAAGTGTCTGCTTTATTTTCTTAAATTTTCTTAAATTTTCTTCCCTCCACCCG 194  
|||||

QY 4535 RACTGCATACCAATTTCTGCAGAGGAGAGTACAGTGCATCTGTAGCTATCTATCTATCTTAC 4594  
|||||

Db 195 ACATTCAAGT---TCTTGGAAACAAGCCATTTTGAACATTTATGTAGTATCTTCTTCCCTAC 251  
|||||

QY 4595 CTACAAATTTGTCTCTCTCTTCCAGGAATATTTGGGATCTGTGAAGGCCCTTCTCTCCAGA 4654  
|||||

Db 252 CAACAACCTCATTTGGCTTTTTCAGAAAAATAGTGAGATAATGAATTCCTACTGATGAAA 311  
|||||

QY 4655 GCGGTGTGCAGATCTGCTTTACTAACTGTTTGAAGCAAAAAACACCGTGGTCAGGT 4714  
|||||

Db 312 GTTGTGCGGTAAACAGTGAAGTTAAAGATATCTAGAAGTGAAGAGATGCTATAGAT 371  
|||||

QY 4715 ACCCTAGAAAAAAGAAATAGTTTGTAGAGCTTCTCTGATGACTATAGCTGCTCTCTGAATC 4774  
|||||

Db 372 ATCCAAGAGAAATCTAACAAATTAATAAACCTTCCAGAGGATTACAGCAGCTCATTAATC 431  
|||||

QY 4775 AAGCTTCTCATTTTCAGGTGCCCCACGGTCTGCAGATCATGAGCGAAAGCATCTCTCTCT 4834  
|||||

Db 432 AAGCATCCAATTTCTCGTGGCCGAAATCAGGTGGTGTATGAAGCAGAGCGGCCCAACTCTGT 491  
|||||

QY 4835 GCCTTTTCTGTGGGCTATACTATGTTCTCAGAAACATTTGCTGCCAGAAATTTGTGAACG 4894  
|||||

Db 492 GCCTTGTGTGGGATCTCTGCTGCTCCAGAGTTACTGCTGCCAGACTGAACCTGGAAG 551  
|||||

QY 4895 GGAAGAGGTTGGAGCTTGCATTTTTCACGCATCTTCACTGTGTGAGCGCGAGTCTGCATTT 4954  
|||||

Db 552 GGGAGGATGTAGGAGCTTGCAGAGCTCACACCTACTCTCTGTGCTCTGGAGTGGGCATCT 611  
|||||

QY 4955 TCCTAAAAATCAGAAATGCCGAGTGGTCTCTGGTGAAGGTAAGCCAGGCTGTGCTCT 5014  
|||||

Db 612 TCCTGAGAGTAGGGAATGTACAGGTGCTATTTTGTAGCTGGCAAAACCAAGGCTGTTTT 671  
|||||

QY 5015 ATCCAGCTCTTACTTGGATGAATATGGAGAAACAGACACCTGGCTTGAAGAGGGGCAACC 5074  
|||||

Db 672 ATTCCTCTCTTACCTTGTGACTATGGGAGACCCAGCAGGAGCTCAGACGGGGAAATC 731  
|||||

QY 5075 CCCTTCATTTATCTCTGAGCGGTATCGGAAGCTCCATTTGGTCTGGCAACAACTGCA 5134  
|||||

Db 732 CTTTACATTTATGCAAGAGCGGATTCAGAAGATTCAGAAGCTCTGGCAACAACTGCA 791  
|||||

QY 5135 TTATAGAAGAGATTTCTAGGAGCCCAAGAGACTAATCAGATGTTTATTTGGATTCACTGGC 5194  
|||||

Db 792 TCACAGAGGAAATTTGGCATGTCACAGGAAAGCAATCAGACACTGTTGTCATTGACTGGC 851  
|||||

QY 5195 A 5195  
Db 852 A 852

## RESULT 7

US-10-071-766-7/c  
; Sequence 7, Application US/10071766  
; Publication No. US20020192678A1  
; GENERAL INFORMATION:  
; APPLICANT: Huel-Mei Chen  
; TITLE OF INVENTION: GENES EXPRESSED IN SENESENCE  
; FILE REFERENCE: PA-0043 US  
; CURRENT APPLICATION NUMBER: US/10/071,766  
; CURRENT FILING DATE: 2002-02-07  
; NUMBER OF SEQ ID NOS: 144  
; SOFTWARE: PERL Program  
; SEQ ID NO 7  
; LENGTH: 4573  
; TYPE: DNA  
; ORGANISM: Homo sapiens

```
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: Incyte ID No. US20020192678A1 1330204.22
; NAME/KEY: unsure
; LOCATION: 3145-3168
; OTHER INFORMATION: a, t, c, g, or other
US-10-071-766-7

Query Match
Best Local Similarity 4.2%; Score 218.2; DB 13; Length 4573;
Matches 440; Conservative 0; Mismatches 308; Indels 17; Gaps 2;

QY 4431 TCCTGGCTGGATTTGGTGGCTCACTCAAGATGGCATACCCCTTATCTTCGCTGTC 4490
Db 4535 TCCGGCTGGCATCTGTGGAGAGTGTGAGAGTTCATGCTCCCTGGAAGTGTTC 4476
QY 4491 TGCATTGTTTTCCTACTATTTACCTGGGTAACCTCGCTGAGGAACGATACCAATTC 4550
Db 4475 TGCATTATTTTTCATTACTTAATGGAGTTCCCTCCCAACCCGACATTCAGT---TCC 4419
QY 4551 TGCAGAGGAGTACAGTCACTCTGTAGCTATCTATCTTTACCTACAAATTTGTTCT 4610
Db 4418 TGGAAACAAGCCATTTGAACATTTATGTAGCTATCTTTCCCTACCAACAACCTCATTTG 4359
QY 4611 GCTCTTCCAGGAATATTCGGATACCTGTAAGCCCTTGTCCAGAGCGGTGTGCAGATCC 4670
Db 4358 CCTTTTCAAGAAATAGTAGATAATGAATCACTGATTGAAAGTTGGTCCGCTAACAG 4299
QY 4671 TGCCTTACTAACTGTTTGAAGCAAAACACACCTGGTGCAGTACCCCTAGAAAAAGAA 4730
Db 4298 TGAAGTTAAAGATATCTAGAGGTGAAGAGATGCTATAGATATCAAGAGAATCTAA 4239
QY 4731 TAGTTTGTAGAGCTTCTGTAGTATAGTCTGCTCCTGATCAAGCTTCTCATTTACAG 4790
Db 4238 CAATTAATAACCTTCCAGAGGATTACAGAGCCTCATTAATCAAGCATCCAATTTCTC 4179
QY 4791 GTGCCCGAGTCTGCAGATGATGAGCGAAAGCATCTGTCTCTGCTTTCTGTGGGCG 4850
Db 4178 GT-----GGTGATAAGAGCAGAGAGCCCAACTCTGTGCTTGTGTGGGATC 4133
QY 4851 TATATATGTTCTCAGAACATTTGCTGCCAGAAATTTGAACGGGAGAGGTTGGAGC 4910
Db 4132 TCTGTCTGTCTCCAGAGTACTGCTGCCAGACTGAACGTGGAAGGGAGGATGTAGGAGC 4073
QY 4911 TTGCAGCTTTCACGCACTTCACTGTGGAGCCGGAGTCTGCAATTTTCTTAAATCAGAGA 4970
Db 4072 CTGCACAGCTCACACCTACTCTGTGGCTCTGGAGTGGCACTTCTCCTGAGAGTACGGGA 4013
QY 4971 ATGCCAGTGTCTCGTTGAAGGTAAAGCCAGAGGCTGTGCTATCCAGCTCCTTACTTT 5030
Db 4012 ATGTCAGGTGCTATTTTGTAGCTGGCAAAACCAAGGCTGTTTTTATCTCTCCTTACCT 3953
QY 5031 GGATGAATATGAGAAACAGACCTTGGCTGAGAGGGGCAACCCCTTCATTTATCTCG 5090
Db 3952 TGATGACTATGGGAGACCGACCGAGGACTCAGAGGGGAAATCCTTTTATCTTATGCAA 3893
QY 5091 TGAGCGGTATCGGAAGTCCATTTGCTTGGCAACAACACTGCATTATAGAGAGATTCG 5150
Db 3892 AGAGCGATTCAAGAGATTCAGAACTCTGGCAACACACATGTCACAGAGAAATTCG 3833
QY 5151 TAGGAGCCCAAGACTAATCAGATGTTATTTGGATTCAACTGGCA 5195
Db 3832 ACATGCACAGGAAGCAATCAGACACTGTTGGCATTTGACTGGCA 3788

RESULT 8
US-10-027-632-260533
; Sequence 260533, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; POLYMORPHISMS IN THE HUMAN GENOME
; FILE REFERENCE: 108827.129

US-09-918-995-27470
; Sequence 27470, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FROM VARIOUS CDNA LIBRARIES
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 27470
; LENGTH: 465
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(465)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-27470
```

Query Match	1.2%.	Score 65;	DB 11;	Length 465;	
Best Local Similarity	49.6%;	Pred. No. 4.9e-07;			
Matches 198;	Conservative	0;	Mismatches 195;	Indels 6;	Gaps 1;
QY	3517	TTTGACATTGGAAAGTGGAGAAATATCTTTGGCCCTCTTTGGCAAAATCTGTGTGCAATACTGTG	3576		
Db	69	TATGATCTAGAAAACGGAGAATTCCTTTGGCTTCTTTGTGAATGCTTCAGTAATACTGTT	128		
QY	3577	ATCCCCATTATTCCTTTTGGCAACCTCAAAAGATAAACAGTGAAGATGCAGATGCTCTTGCT	3636		
Db	129	ATTCC-----CTGTGCTGCTCTCTCCCAAGAAATATTTTAAACAACAGGTTAAATTTTCA	182		
QY	3637	CAACTTTTGACCCTGGCAGCGTGGATACAGACTCTCTCTGGCCAGAAATATCAGGTTATAAT	3696		
Db	183	GACCAACCAATCTGACTCAGTGGATTAGACAAATATCTCAGCAAAATAAAGCAATTACAG	242		
QY	3697	ATAAGACATGCTAAAGGAGAAAAACCCCAATTCCTATTTTCTTAATCAAGGAATGGGAGAT	3756		
Db	243	TTTCTTAGGAAAAGAAAGTACTCTCTTAATAATGCCCTACAAAAGAAATCAGAAAATGTG	302		
QY	3757	TCIACATTGGAGTCCCAATCCATCCTCAGTTTTGGCGTTGAGTCTTCGATTAAATATCA	3816		
Db	303	GATGAATACAGCTCCCTGAAGGTTTCAGGCCTGANTTCGTCTAGATGCCCTATTCTCT	362		
QY	3817	AATAGCATCAAGGAAATGGTTATTCTCTTTGCCACAACAATTTATAGAATTGGATTGAAA	3876		
Db	363	GAGAGCATAAAGAAATGCTAACGACATTTGGAACTGCTACCTACAGGTGGGACTAAG	422		
QY	3877	GTGCCACCTGATGAAGAGGATTCCTCAGTCCCCCATGTG	3915		
Db	423	GTTCATCCCAATGAAGAGGATTCCTCGTGTGCCATAATG	461		

RESULT 10  
US-09-908-975-8907  
; Sequence 8907, Application US/09908975  
; Publication No. US20030165843A1  
; GENERAL INFORMATION:  
; APPLICANT: SHOSHAN, Avi  
; APPLICANT: WASSERMAN, Alon  
; APPLICANT: MINTZ, Eli  
; APPLICANT: MINTZ, Liat  
; APPLICANT: FAIGLER, Simchon  
; TITLE OF INVENTION: OLIGONUCLEOTIDE LIBRARY FOR DETECTING RNA TRANSCRIPTS AND SPLICE  
; TITLE OF INVENTION: THAT POPULATE A TRANSCRIPTOME  
; FILE REFERENCE: 36688-0005  
; CURRENT APPLICATION NUMBER: US/09/908,975  
; CURRENT FILING DATE: 2001-07-20  
; PRIOR APPLICATION NUMBER: US 60/287,724  
; PRIOR FILING DATE: 2001-05-02  
; PRIOR APPLICATION NUMBER: US 60/221,607  
; PRIOR FILING DATE: 2000-07-28  
; NUMBER OF SEQ ID NOS: 32337  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 8907  
; LENGTH: 60  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-908-975-8907

	Query Match	1.2%	Score 60	DB 12	Length 60
	Best Local Similarity	100.0%	Pred. No. 2.7e-06		
	Matches 60	Conservative 0	Mismatches 0	Indels 0	Gaps 0
QY	2199	CTATTGTGGGTGAGCGTTATGTACCTGGAGTGGGAAATGTGACCAAAAGAGGTCAC	2258		
Db	1	CTATTGTGGGTGAGCGTTATGTACCTGGAGTGGGAAATGTGACCAAAAGAGGTCAC	60		

RESULT 11  
US-10-027-632-8115  
; sequence 8115, Application US/10027632

```

; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8115
; LENGTH: 611
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(611)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-8115

```

Query Match	1.0%	Score 53.8	DB 13	Length 611
Best Local Similarity	96.3%	Pred. No. 0.00074		
Matches 55	Conservative 0	Mismatches 2	Indels 0	Gaps 0
QY	3900	TCGAGTCCCATGCTGACCTGGAGCAGCTCGCGTTTTCACATATCAGGCAATGAAA	3956	
Db	1	TCGAGTCCCATGCTGACCTGGAGCAGCTCGCGTTTTCACATATCAGGCAATGGTAA	57	

```

RESULT 12
US-10-171-581-278
; Sequence 278, Application US/10171581
; Publication No. US2003010426A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: Linsley, Peter
; APPLICANT: Mao Mao
; TITLE OF INVENTION: Signature Genes in Chronic Myelogenous Leukemia
; FILE REFERENCE: 9301-157-999
; CURRENT APPLICATION NUMBER: US/10/171,581
; CURRENT FILING DATE: 2002-06-14
; PRIOR APPLICATION NUMBER: 60/298,914
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 366
; SEQ ID NO 278
; LENGTH: 535
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: Modified_base
; LOCATION: 1 ... 535
; OTHER INFORMATION: n = a,c,g, or t
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: N38966
; DATABASE ENTRY DATE: 2001-06-18
US-10-171-581-278

```

Query Match	0.8%	Score 44.2;	DB 14;	Length 535;
Best Local Similarity	63.8%	Pred. No. 0.31;		
Matches 67;	Conservative	0;	Mismatches 38;	Indels 0;
Matches 0;	Gaps	0;		

RESULT 15  
US-09-814-353-21640  
; Sequence 21640, Application US/09814353  
; Publication No. US20030165831A1  
; GENERAL INFORMATION:  
; APPLICANT: Lee, John  
; APPLICANT: Thompson, Pamela  
; APPLICANT: Lillie, James  
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR  
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND  
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER  
; FILE REFERENCE: MRI-006B  
; CURRENT APPLICATION NUMBER: US/09/814,353  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: US 60/191,031  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: US 60/207,124  
; PRIOR FILING DATE: 2000-05-25  
; PRIOR APPLICATION NUMBER: US 60/211,940  
; PRIOR FILING DATE: 2000-06-15  
; PRIOR APPLICATION NUMBER: US 60/216,820  
; PRIOR FILING DATE: 2000-07-07  
; PRIOR APPLICATION NUMBER: US 60/220,661  
; PRIOR FILING DATE: 2000-07-25  
; PRIOR APPLICATION NUMBER: US 60/257,672  
; PRIOR FILING DATE: 2000-12-21  
; NUMBER OF SEQ ID NOS: 22037  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 21640  
; LENGTH: 2372  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:

; NAME/KEY: misc\_feature  
; LOCATION: 2372  
; OTHER INFORMATION: n = A,T,C or G  
US-09-814-353-21640

Query Match 0.8%; Score 43.8; DB 12; Length 2372;  
Best Local Similarity 53.0%; Pred. No. 1.1;  
Matches 149; Conservative 0; Mismatches 117; Indels 15; Gaps 2;

QY	4819	AAGCATGCTGCTCTGCGCTTTTCTGTGGGGCTATACTATGTCTCAGAACATTGTGCTGC	4878
Db	990	AAAGATCCTGCTGTTGCGCTTGTGTGGTACTTTTCTATGCTGAAAGGACTTTGCTGC	1049
QY	4879	CAGGAATTTGTGAACGGGGAAGAGTTGGAGCTTGCATTTTTCACGCACATTCACGTGGA	4938
Db	1050	AAGCAACAAAGTTACTGTGAA-----TGTGTACTGCACTCTCAGAACTGTGGT	1097
QY	4939	GCCGGAGTCTGCAATTTTCCTAAATAATCAGAGATGCCGAGTGGTCTGGTTGAAGGTAAA	4998
Db	1098	GCAGGAACAGGTATTTTCCTTTTGTGATCAATGCATCGGTAATTATCATCATTCGAGGTCA-	1156
QY	4999	GCAGAGGCTGTGCTATCCAGTCCCTTACTTGGATGAATATGGAGAAACAGACCCTGGC	5058
Db	1157	--CCGCTTCTGCGCTCTGGGGTTCGGTGTATTGGATGCTCATGGAGAGAGACCGGGAT	1214
QY	5059	CTGAAGAGGGGCAACCCCTTCATTTATCTCGTGAGCGGTA	5099
Db	1215	CTTAGCGAGGCAACCTCTACATTTGTAAGGAAAGATA	1255

Search completed: September 27, 2003, 20:27:15  
Job time : 823.228 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 27, 2003, 00:42:14 ; Search time 6516.54 Seconds  
(without alignments)  
19412.893 Million cell updates/sec

Title: US-09-724-126a-18  
Perfect score: 5205  
Sequence: 1 atggcgagagagagctgg.....tcaactggcagttactgtga 5205

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_estc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_est3:\*  
12: gb\_est4:\*  
13: gb\_est5:\*  
14: gb\_est6:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rod:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vrl:\*  
28: gb\_gss1:\*  
29: gb\_gss2:\*

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1981.6	38.1	2561	11 AK089616	AK089616 Mus muscu
2	825	15.9	960	12 BM472160	BM472160 AGENCOURT
3	804.6	15.5	3641	11 BC044903	BC044903 Mus muscu
4	771.4	14.8	865	13 BQ233617	BQ233617 AGENCOURT

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

5	708.8	13.6	712	13	BX104087	BX104087
6	673	12.9	682	9	AI929033	AI929033
c	664.8	12.8	668	13	B0618516	B0618516
8	656.6	12.6	756	10	BG534574	BG534574
9	628.6	12.1	3627	11	AK083320	AK083320
10	620	11.9	797	12	BG862813	BG862813
11	596	11.5	774	13	B0955331	B0955331
12	585.2	11.2	695	14	B0756516	B0756516
c	583.6	11.2	703	12	BM114217	BM114217
13	577.4	11.1	3058	11	AK078173	AK078173
15	576	11.1	729	13	B0703363	B0703363
16	571.6	11.0	782	12	B1086469	B1086469
17	560	10.8	560	14	CB159971	CB159971
18	545.4	10.5	684	14	B0764809	B0764809
19	521.2	10.0	528	13	BX119443	BX119443
20	517.6	9.9	691	14	B0736138	B0736138
21	512.2	9.8	606	10	BE589438	BE589438
c	511.4	9.8	752	14	CA313244	CA313244
23	507.8	9.8	633	10	BB622408	BB622408
24	507.2	9.7	659	14	BY750548	BY750548
c	497.4	9.6	719	13	B0623096	B0623096
26	494.4	9.5	496	13	BX281065	BX281065
27	494	9.5	495	14	CB129142	CB129142
c	487.2	9.4	1105	13	B0751567	B0751567
c	485.4	9.3	641	9	AI361043	AI361043
c	482.4	9.3	506	2	HSN077106	HSN077106
c	479.8	9.2	624	10	BG219270	BG219270
c	477.8	9.2	481	9	AI192195	AI192195
c	473.6	9.1	609	29	CC200266	CC200266
34	473.6	9.1	777	14	CB232501	CB232501
35	470	9.0	478	9	AA401319	AA401319
36	453.8	8.7	737	13	B0461940	B0461940
c	451.2	8.7	465	12	BM508126	BM508126
38	447.4	8.6	522	10	BG382648	BG382648
39	446.6	8.6	830	14	CB588645	CB588645
40	446.4	8.6	520	10	BG382624	BG382624
c	446.4	8.6	640	14	CD368666	CD368666
42	445.2	8.6	710	14	CB519911	CB519911
43	441.2	8.5	583	12	BI540789	BI540789
44	440	8.5	751	13	B0269526	B0269526
45	439.8	8.4	905	13	B0272014	B0272014

## ALIGNMENTS

RESULT 1  
AK089616  
LOCUS  
DEFINITION  
AK089616 2561 bp mRNA linear HTC 05-DEC-2002  
Mus musculus activated spleen cDNA, RIKEN full-length enriched  
library, clone:F830005C07 product:ubiquitin protein ligase E3  
component n-recognin 1, full insert sequence.  
ACCESSION  
AK089616  
VERSION  
AK089616.1 GI:26354612  
KEYWORDS  
HTC; CAP trapper.  
SOURCE  
Mus musculus (house mouse)  
ORGANISM  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1  
Carninci, P. and Hayashizaki, Y.  
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new genes  
Genome Res. 10 (10), 1617-1630 (2000)  
20499374  
PUBMED  
11042159



REFERENCE  
AUTHORS

3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsunai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Wataniki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y. RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer  
Genome Res. 10 (11), 1757-1771 (2000)  
20530913  
11076861

REFERENCE  
AUTHORS

4  
Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,  
Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S.,  
Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I.,  
Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R.,  
Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T.,  
Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H.,  
Kuehl, P., Lewis, S., Matsuo, Y., Nikaudo, I., Pesole, G.,  
Quackenbush, J., Schriml, J. M., Staubli, F., Suzuki, R., Tomita, M.,  
Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H.,  
Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N.,  
Carinci, P., de Bonaldo, M. F., Brownstein, M. J., Bult, C.,  
Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D.,  
Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P.,  
Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P.,  
Ring, B., Rongwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,  
Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H.,  
Toyooka, K., Wang, K. H., Weitz, C., Whittaker, C., Wilming, L.,  
Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohtsuki, S.,  
and Hayashizaki, Y.  
Functional annotation of a full-length mouse cDNA collection  
Nature 409 (6821), 685-690 (2001)  
21085560  
11217851

REFERENCE  
AUTHORS

5 The FANTOM Consortium and the RIKEN Genome Exploration Research  
Group Phase I & II Team.  
Analysis of the mouse transcriptome based on functional annotation  
of 60,770 full-length cDNAs  
Nature 420, 563-573 (2002)  
6 (bases 1 to 2561)

JOURNAL  
REFERENCE

Nature 420, 563-573 (2002)  
6 (bases 1 to 2561)  
Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carrincci,P., Fukaya,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hamamoto,K., Hiraoka,T., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T., Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohmoto,N., Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sugabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Takeda,Y., Tanaka,T., Tomaru,A., Toyota,T., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.

TITLE  
JOURNAL

Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp). URL: <http://genome.gsc.riken.go.jp/>, Tel: 81-45-503-9222, Fax: 81-45-503-9216.

COMMENT

CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Tissues were provided by Dr. John Todd (Dept. of Medical Genetics Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome Trust/MRC building Addenbrookes Hospital Cambridge) whose assistance we gratefully acknowledge.

Please visit our web site for further details.  
 URL: <http://genome.gsc.riken.go.jp/>  
 URL: <http://fantom.gsc.riken.go.jp/>.

FEATURES	SOURCE
----------	--------

```

1. 2561
/organism="Mus musculus"
/mol_type="mRNA"
/strain="NOD"
/db_xref="FANTOM_DB:F830005C07"
/db_xref="taxon:10090"
/clone="F830005C07"
/tissue_type="activated spleen"
/clone.lib="RIKEN full-length enriched mouse cDNA library"
14. >2560

```

CDS

/note="unnamed protein product; putative  
 ubiquitin protein ligase E3 component 1  
 (MGDI|GI:1277977, GB|NM\_009461, evidence: BLASTN, 100%,  
 match=2560)"  
 /codon\_start=1  
 /protein\_id="BAC40933.1"  
 /db\_xref="GI:26354613"  
 /translation="MADEMDGASLRMDVSPPEPLAPQRPASWMDQQVDYAFALHLHLA  
 OLVPETYFAEMDPDLSEKQERSVQMSILTPLEWYLEGEDPDICLEKLEKHSAGAFCEKV  
 FKSGETYSRCDAIDPTCVICMDCFQSSVHKHNRKYMHTSTGGGFCDCGDTAEWTKG  
 PFCVDHPGRAGTKTDLKNEEVYEAQARRIPPSVYKIVEMTEWEEKELPPPELQ  
 IRKRENYCYVLEDESHSDYHVIYSLQADCELAEPALQHTTAMDKEGRRAVKAGVY  
 QCOAEKEDIKSENVPQHPLEVYHSHVMAHQKRALFGSWMMKINMSYSDFFRQI  
 FTCAQCLVEEPPSNPCLISRLMLDARKLYGARKILHLEIFSSFMEMEYKKLFAMEF  
 VYKYKOLQKYISDDHSESTIALSVMLTPTLRAHLIEQONVTSVETLEVLV  
 EYLDRNNKFGCYSDSKLGRVAVTCMLYILISKPVWTERLAQFLGRRPSFLKI  
 LTCMQCEETRRQGVGHVDPDWEARAATOMOLKNILLMFQEWCADEDLLVAAYKE  
 CHKAVNRCSSTNFMSSYVQLGCHLSLETASYKVSEDLVSIHLPLSRLTALVHLVSR  
 LGLAISLHEVPSPDSQVEVLVETPLCLVLAQVVAEMRRNRGLSLISQVYFYQDVK  
 CREEMVDKDIILQIGASIMDPNKRFLLLVQRYVELDAPNKTISTDKDDILQOYNTLI  
 EEMLVLIYTVGRGYVPGSNVDEEYVREIMREITHLICIEPMHSAARNLPPENNET  
 GLENVINKVATFKKPGYSGHGVLEKDESILKDFNMVYFYHSYKQHSKAEHM"  
 763 a 509 c 617 q 672 t

[illegible]

Query Match	38.1%	Score 1981.6	DB 11	Length 2561
Best Local Similarity	86.5%	Pred. No. 0		
Matches 2205	Conservative 9	Mismatches 289	Indels 45	Gaps 1
Qy	1	ATGCGCACGAGGAGCGTGGAGGTACTGAGAGATGAAATCAGCGCGGAGTTACCCAG	60	
Db	14	ATGCGGACGAAGAGATGGAGCGGCCGAGAGATGGAGCTGACCCCGGAGCCTCCCTG	73	
Qy	61	ACCCTTCAGGCGTCTGGCATCTTGGTGGGATCAGCAAGTTGATTTTATATCTGTTCTTG	120	
Db	74	GCCCGCAGCGCGCGCATCGTGTGGGATCAGCAAGTTGATTTCTATACTGCTTCTTA	133	
Qy	121	CATCATTTGGCACAATTTGGTGGCCAGAATTTACTTTGCTGAAATGGACCCAGACTGGAA	180	
Db	134	CATCATTTGGCACAATTTAGTGGCCAGAATTTATTTGCTGAGATGGACCCAGATTTGGAA	193	
Qy	181	AAGCAGGAGAAAGTGTACAAATGTCAAATTTCACTCCACTGGAATGGTACTTATTTGGA	240	
Db	194	AAGCAAGAGAGAGTGTACAGATGTCAACTCACTCTTTGGAGTGGTACTTATTTGGA	253	
Qy	241	GAAGATCCAGATATTTGCTTTAGAGAAATTTGAACACAGTGGAGCATTTACGCTTTGTGGG	300	
Db	254	GAGGATCCGGATATTTGCTTTAGAGAAATTTAAACACACAGTGGAGGTTTCCAGTTGTGTGGG	313	
Qy	301	AGGTTTTTCAAAAGTGGAGAGACAACTATTTCTTGACGGATTTGTGCAATTTGATCCAACA	360	
Db	314	AAGTTTTTCAAAAGTGGAGAAACAATATTTCTGTAGGGATTTGTGCATTTGATCCAACG	373	
Qy	361	TGTTACTCTGTATGGAGTCTTCCAGGACAGTGTTCATAAAATCATCTGTACAAGATG	420	
Db	374	TGTTGCTCTGTATGGACTGTTCCAAAGTAGTGTTCATAAAAACCATCTGTACAAGATG	433	
Qy	421	CATACTTCTACTGGAGGAGGTTCTGTGACTGTGGAGACACAGAGGCATGGAATCTGGC	480	

Db 434 CATACTTCTACTGAGGGGCTTCTGTGACTGTGGAGACACAGAAGCGTGGAAACTGGC 493  
Qy 481 CCTTTTGTGTAAATCATGAACCTGGAGAGACAGGTACTATAAAGAGAAATTCAGCGTGT 540  
Db 494 CCTTTTGTGTGGATCAGGACCTGGAAAGACAGGTACTACAAAGAGAGAGCTTACATTCG 553  
Qy 541 CCGTTGAATGAAGAGGTAAATGTCTCAAGCCAGGAAATATTTCTTCAGTCATATAAATAT 600  
Db 554 CCATTGAATGAAGAGGTGATGCTCAAGCCAGGAGAAATATTCCTTCGGTATATAAATAC 613  
Qy 601 GTCGTAGAAATGACTATATGGGAAGAGGAAAGAACTGCTCTGAACTCCAGATAAGG 660  
Db 614 ATTGTAGAAATGACTATATGGGAAGAGGAAAGAAATTCCTCTGAACTGCAGATAAGG 673  
Qy 661 KNR-----YVCVNDHSHSDH 675  
Db 674 GAGAAAAATGAACGATACTATTGTGCTCTTTCAACGATGAGCCACTTCGTATGATCAT 733  
Qy 676 GTCATATACAGCCTACAAAGAGCTTTGACTGTGAGCTCGCAGAGGCCAGTTCGATACC 735  
Db 734 GTGATCTACAGTCTGCAGAGAGCTCTAGATTGCGAGCTTGCAGAGGCACAGCTGCACAG 793  
Qy 736 ACTGCCATTGACAAAGAGGCTGTCGGGCTGTTAAAGCGGAGCTTTATGCTGCTTGCAG 795  
Db 794 ACTGCCATCGACAAAGAGGCTGCGGGCTGTCAAGCAGAGTGTGATGCCACTTGCAG 853  
Qy 796 GAAGCAAAGGAGATATAAGAGTCAATTCAGAAATGTCTCTCAACATCCACTTCATGTA 855  
Db 854 GAAGCAAAGGAGGATATAAGAGTCACTCAGAGAACGTCTCTCAGACCCCTCCATGTG 913  
Qy 856 GAAGTATTACACTCAGAGATTATGGCTCATCAGAAATTTGCTTTCGGCTTTGGTCTCTGG 915  
Db 914 GAGTGTGTCACCTCGGTGTTATGGCTCACAGAAATTCGCTCTGCGGCTTTGGTCTCTGG 973  
Qy 916 ATGAACAAATATGAGCTATTCAAGTGACTTTAGGCAGATCTTTTGGCAAGCATGCCCT 975  
Db 974 ATGACAAATATGAGCTATTCAAGTGACTTTAGACAGATATTTTGGCCAGGCTCGCTC 1033  
Qy 976 AGAGAAGACCTGACTCGGAGAAATCCCTGTCTCATAAGCAGGTTAAATGCTTTGGGATGCA 1035  
Db 1034 GTAGAAGAACCTGGCTCTGAAATCCCTGTCTTATAAGCAGACTAAATGCTTTGGGATGCA 1093  
Qy 1036 AAGCTTTATAAGGTGCGCTTAAGATCCTTCATGAATGATCTTCAGCAGTTTTTTTATG 1095  
Db 1094 AAAGCTTTATAAGGTGCGCTTAAGATCCTTCATGAATGATCTTTAGTAGTTTTTTTATG 1153  
Qy 1096 GAGATGAATACAAAAAATCTTTGCTATGGAATTTGTGAAGTATTTATAACAACATGCGAG 1155  
Db 1154 GAGATGAATACAAAAAATCTTTGCTATGGAATTTGTGAAGTATTTATAACAACATGCGAG 1213  
Qy 1156 AAAGAATATATCAGTGATGATCATGACAGAGTATCTCTATAACTGCACTTTTCAGTTTCAG 1215  
Db 1214 AAAGATFACATCAGCAGCAGCAGCAGAGAGAGCATCTCCATAAAGCGGCTGTCGTCGAG 1273  
Qy 1216 ATGTTTACTGTTCTCTACTGCTGCGATCTTATTTGAAGAGAGATGTTTATCTCTGTC 1275  
Db 1274 ATGCTCACCTGCGGCTTGGCCGCGCATCTTATTGAAGAGCAGATGTTATTCTCTGTC 1333  
Qy 1276 ATTACTGAACTCTGCTAGAGTTTTTACCTGAGTACTTGGACAGGAAATATAATTCAC 1335  
Db 1334 ATTACTGAAAGCTGCTAGAGTTTTTACCTGAATCTTGGACAGGAAATATAATTCAC 1393  
Qy 1336 TTCCAGGTTTATAGCCAGGACAAATTTGGGAAGAGTATATGAGTAATATGACCTTAAG 1395  
Db 1394 TTCCAGGTTTATAGCCAGGACAAATCTGGGAAGAGTCTACGCGATTATATGACCTTAAG 1453  
Qy 1396 TATATCTGATCAGCAACCCACAAATATGGACAGAAAGATTAAGAAATGCACTTCTTGA 1455  
Db 1454 TATATCTGATACAGGCTGTCTATATGGACAGACGATTAAGAGCGCAGTTCTCTGGAA 1513  
Qy 1456 GGTTTTCGATCTTTTGAAGATCTTACCTGTATGAGGGAATGGAAGAAATCCGAAGA 1515  
Db 1514 GGTTCGCGCTTTTCTGAAGATCTTACCTGTATGAGGGAATGGAAGAAATCAGAAGA 1573

RESULT 2  
BM472160  
LOCUS

BM472160

960 bp

mrna

linear

EST 05-FEB-2002

Qy 1516 ~CAGGTTGGCCACACACATTGAAAGTGGATCTCTGATTGGAGGCTGCCAATTTGCTATACAGATG 1575  
Db 1574 CAAGTTGGCAACACACATTGAAAGTGGAGCTGAGTGGAGGCTGCCATTCGCTATACAGATG 1633  
Qy 1576 CAATTTGAAGAATATTTTACTCATGTTTCCAAAGAGTGTGCTTGTGATGAAGAACTCTTA 1635  
Db 1634 CAACTAAAGAATATTTTGTCTCATGTTCCAAAGAGTGTGCTTGTGATGAAGATCTCTTA 1693  
Qy 1636 CTTGTGGCTTATAAAGAATGTCCAAAGCTGTGATGAGGTGCGAGTACCAGTTTTCATCT 1695  
Db 1694 CTGTGGCTTATAAAGAATGTCCAAAGCTGTAATCAGGTGCGAGTACAAATTTTCATGTCC 1753  
Qy 1696 AGTACCAACAGACAGTAGTACAATCGTGTGCACATAGTTTGGAAACAAGTCCCTACAGATA 1755  
Db 1754 AGTACCAACAGACAGTAGTCAATTTGCGGTCTAGTCTGAAACCAATCTCAAAAGTG 1813  
Qy 1756 TCTGAGGATCTTTTAAGCATACATCTGCCACTCTCTAGGACCCCTGCTGCTCTTCATGTA 1815  
Db 1814 TCTGAGGACCTTGTAAAGCATACCTGCCACTCTCTAGAACACTTCTGCTCTTCATGTA 1873  
Qy 1816 CTTTTAAGCAGGCTGGTCTCTTTCAAGACTGCATGAATTTGTCTTTTGGAGACTTT 1875  
Db 1874 CGTTTTAAGCAGACTAGGTCTTATTTCAAGACTGCATGAATTTGCTGCTTTTGACAGCTTT 1933  
Qy 1876 CAAGTAGAGTACTAGTGAATATCTTTTACGTTGTCTGCTGTTGGTGGTCCAGGTTGTT 1935  
Db 1934 CAAGTAGAGTCTGCTGGTGAATACCGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1993  
Qy 1936 GCTGAGATCTGCGGAAGAATAGCTGTCTCTTATTAGCCAGTGTGTTTATTACCAGAT 1995  
Db 1934 GCTGAGATCTGCGGAAGAATAGCTGTCTCTCTCTCCTCCTCAGCAGGTTTTTCTATTATCAGAT 2053  
Qy 1936 GTTAAGTGCAGAGAAATGATGATAAAGATATCATCTGCTTCAGATTCAGATTTGGTGCATCT 2055  
Db 2054 GTTAATGCAGGAGGAAATGTACGATAAAGATATCATCTGCTTCAGATTTGGAGCATCT 2113  
Qy 2056 TTAATGGATCCCAATTAAGTCTTGTGTTACTTGGTACTTTCAGAGTATGAATTTGGCGAGGCT 2115  
Db 2114 ATAATGGATCCCAACAAGTCTTGTGTTACTTGGTACTTTCAGAGATATGAATTTACTGATGCT 2173  
Qy 2116 TTTAACAGACCATATCTACAAAGACCAGGATTTGATTAACAATATAATACATAATA 2175  
Db 2174 TTTAACAGACCATATCTCCAAAGACCAGGATTTGATTAACAAGTATAATACATAATA 2233  
Qy 2176 GAAGAAATGCTTTCAGGCTCCTCATCTATATTGTTGGTGGTGGTGTATGCTGAGGAGTGGGA 2235  
Db 2234 GAAGAAATGCTTTCAGGCTCCTCATCTATATTGTTGGGAGAGCTTATGCTGAGTGGGA 2293  
Qy 2236 AATGTACCAAAAGAGAGTCAATGAGAGAAATCATTCACTTGCTTTGCTGATTTGAACCC 2295  
Db 2294 AATGTTACCAGAGAGAGGTTTAAATGAGAGAGATTTACTCATTCTTTCATTTGAGCCC 2353  
Qy 2296 ATGCCACAGTGGCATTTGCCAAAATTTACCTGAGAAATGAAATATGAACTGGCTTA 2355  
Db 2354 ATGCCACAGTGGCATTTGCCAGAAACCTACCTGAGAACGAAATATGAACTGGCTTA 2413  
Qy 2356 GAGAATGTCATAAACAAGTGCCACATTTAAGAAACCAGGTGTATCAGGCCATGGAGTT 2415  
Db 2414 GAGAATGTCATAAACAAGTGCCACATTTAAGAAACCAGGTGTGTCGGGCCATGGAGTT 2473  
Qy 2416 TATGAACATAAAGATGAATCACTGAAAGACTTCAATATGATGCTTTTATCTACTTACTCCAAA 2475  
Db 2474 TATGAATGAAGATGAATCACTGAAAGACTTCAATATGATGCTTTTACCATTATTCTAAA 2533  
Qy 2476 ACCCAGCATPAGAGGCTGAACATATGC 2503  
Db 2534 ACACAGCATPAGAGGCTGAACATATGC 2561

DEFINITION AGENCOURT\_6469668 NIH\_MGC\_92 Homo sapiens cDNA clone IMAGE:5576383  
5', mRNA sequence.  
ACCESSION BM472160  
VERSION BM472160.1 GI:18521202  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (Bases 1 to 960)  
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: ATCC  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLAM12328 row: c column: 08  
High quality sequence stop: 720.  
FEATURES  
source  
1..960  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5576383"  
/tissue\_type="embryonal carcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_92"  
/note="Organ: testis; Vector: pCMV-SPORT6; Site\_1: NotI;  
Site\_2: SalI; Cloned unidirectionally; oligo-dT primed.  
Average insert size 2.5 kb. Library enriched for  
full-length clones and constructed by Life Technologies.  
Note: this is a NIH\_MGC library."  
BASE COUNT 230 a 227 c 218 g 283 t 2 others  
ORIGIN  
Query Match 15.9%; Score 825; DB 12; Length 960;  
Best Local Similarity 96.0%; Pred. No. 1e-158;  
Matches 889; Conservative 0; Mismatches 31; Indels 6; Gaps 4;  
QY 4184 TGTAGCATCCCACTCTGATGGGATGAGCCCTGTTGATCGAGCGCTTCTTCAGTTA 4243  
DB 1 TGTAGCATCCCACTCTGATGGGATGAGCCCTGTTGATCGAGCGCTTCTTCAGTTA 60  
QY 4244 GTTCTCTATAAACACCTTTATCTCTCCATTGATCACCATGGCACACATGCTTCAGA 4303  
DB 61 GTTCTCTATAAACACCTTTATCTCTCCATTGATCACCATGGCACACATGCTTCAGA 120  
QY 4304 TACTACTTACAGTACACAGCGCTTACCCCTGCTCAGGTTCAAGAAGACAGTGAAGAG 4363  
DB 121 TACTACTTACAGTACACAGCGCTTACCCCTGCTCAGGTTCAAGAAGACAGTGAAGAG 180  
QY 4364 CTCATTCGCGATCTCTTTCTTTGAGAAATTTCTCAATATACAAGTGGCTCCATTGGGT 4423  
DB 181 CTCATTCGCGATCTCTTTCTTTGAGAAATTTCTCAATATACAAGTGGCTCCATTGGGT 240  
QY 4424 GTGATATTCCGGCTGGTATTGTTGGGTCTCACTGAAGAATGGCATCCCCCTTATCTTC 4483  
DB 241 GTGATATTCCGGCTGGTATTGTTGGGTCTCACTGAAGAATGGCATCCCCCTTATCTTC 300  
QY 4484 GCTGTGCTGCATTTGTTTTCACATATTTACTTGGGGTAACTCCGGCTGAGGAACATGCATA 4543  
DB 301 GCTGTGCTGCATTTGTTTTCACATATTTACTTGGGGTAACTCCGGCTGAGGAACATGCATA 360  
QY 4544 CCAATTCTGCAGAAGGAGATACAGTGCACCTCTGTAGCTATCTATCTTTACCTACAAATT 4603  
DB 361 CCAATTCTGCAGAAGGAGATACAGTGCACCTCTGTAGCTATCTATCTTTACCTACAAATT 420

QY 4604 TGTTCCTGCTCTTCCAGGAATATTGGATACTCTAAGGCCCTTGCCTCCAGAGCGGTGTG 4663  
DB 421 TGTTCCTGCTCTTCCAGGAATATTGGATACTCTAAGGCCCTTGCCTCCAGAGCGGTGTG 480  
QY 4664 CAGATCCTGCTCTTACTAAACTGTTTGAAGCAAAAACACCGTGGTCCAGGTACCTAGAA 4723  
DB 481 CAGATCCTGCTCTTACTAAACTGTTTGAAGCAAAAACACCGTGGTCCAGGTACCTAGAA 540  
QY 4724 AAAGAATAAGTTTGTATAGAGCTTCCCTGATGACTATAGTCCCTCCTGAATCAAGCTTCTC 4783  
DB 541 AAAGAATAAGTTTGTATAGAGCTTCCCTGATGACTATAGTCCCTCCTGAATCAAGCTTCTC 600  
QY 4784 ATTTCAGGTGCCACCGTCTGCAGATGATGAGCAAAAGCATCTGCTCTGCTTTTCT 4843  
DB 601 ATTTCAGGTGCCACCGTCTGCAGATGATGAGCAAAAGCATCTGCTCTGCTTTTCT 660  
QY 4844 GTGGGCTATATCTATCTCTCAGCAATTTGCTGCCAGGAAATTTGTAACGGGGAACAGG 4903  
DB 661 GTGGGCTATATCTATCTCTCAGCAATTTGCTGCCAGGAAATTTGTAACGGGGAACAGG 720  
QY 4904 TTGGAGCTTGCATTTTTCACGCACTTCTCAGTGTGGAGCCGGAGTCTGCATTTTCTTAAAA 4963  
DB 721 TTGGAGCTTGCATTTTTCACGCACTTCTCAGTGTGGAGCCGGAGTCTGCATTTTCTTAAAA 780  
QY 4964 TCAGAGAATGCCGAGTGG-TCCTGGTTGAAGGTAAGCCAGAGGCTGT-CCCTATCCAGC 5021  
DB 781 TCAGAGAATGCCGAGTGGTTCCTGGTTGAAGGTAAGCCAGAGGCTGTCCCTATCCAGC 840  
QY 5022 T-CCTTACTTGGATGAATATGGAGAA--ACAGACCTCGCTGAGAGGGGGAACCC 5077  
DB 841 TCCCTTACTTGGATGAATATGGAGAAACAGAACCTTGGCCCTGAAGAGGGGGAACCC 900  
QY 5078 TTCAATTTATCTCGTGAGCGGTATCGG 5103  
DB 901 CCCTTTCAATTTCTCCTGGAGCGG 926  
RESULT 3  
BC044903  
LOCUS  
DEFINITION BC044903 3641 bp mRNA linear HTC 10-FEB-2003  
Mus musculus, Similar to ubiquitin ligase E3 alpha-II, clone  
IMAGE:3493115, mRNA.  
ACCESSION BC044903  
VERSION BC044903.1 GI:28279967  
KEYWORDS HTC.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
Strausberg, R.  
Direct Submission  
Submitted (23-JAN-2003) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: Gilbert Smith, Ph.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Baylor College of Medicine Human Genome  
Sequencing Center  
Center code: BCM-HGSC  
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>  
Contact: [ang@bcm.tmc.edu](mailto:ang@bcm.tmc.edu)  
Gunnaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Lounseged, H.,  
Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,  
A.N., Gibbs, R.A.  
Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Series: IRAK Plate: 40 Row: b Column: 10  
This clone has the following problem: frame shifted.

**FEATURES**

```

FEATURES
source
1. .3641
/location/qualifiers
/organism="Mus musculus"
/mol_type="mRNA"
/strain="CZECH II"
/db_xref="taxon:10090"
/clone="IMAGE:3493115"
/tissue_type="Mammary tumor"
arose spontaneously from a
(clonal) outgrowth infected
(clone_lib="NCI CGAP Lu29"
/lab_host="DH10B"
/notes="Vector: pCMV-SPORT6"
989 a 9899 c 906 g 847 t
BASE COUNT

```

Query Match	15.5%	Score	804.6;	DB	11;	Length	3641;
Best Local Similarity	54.7%;	Pred.	No. 1.9e-154;				
Matches	1928;	Conservative	0;	Mismatches	1509;	Indels	88;
Gaps	13;						
QY	1720	TGTGGACATAGTTTGAAACAAAGTCCTCAGAGTATCTCAGGATCTTGTAAAGCATACAT	1779				
Db	22	TGTGGACATCGGTGGAAACCATCAGATACTGTGTTCCCAAGAAAGATGTAGCATTCAC	81				
QY	1780	CTGCCACTCTAGGACCCCTGTGCTCTCTCATGTACGTTTAAGCAGGCTGGGTGCTGTT	1839				
Db	82	CTCCCAATTCTCGCTTGCAGCTTTGCATGTATTGTTAAGCAAAAGCTGAAGTGGCA	141				
QY	1840	TCAAGACTCGATGAATTTGTGCTTTTGGAGACTTTCAGTAGAGGTACTAGTGGAAATAT	1999				
Db	142	TATAAATTTCCAGAGCTCTACCTCTAAGTGAACCTGAGCCACCCATGTTGATAGAGCAT	201				
QY	1900	CTTTTACGTTGCTGCTGTGTTGGTTCGCCAGGTGTTGCTCAGATGTGGCGCAAGAAATGGA	1959				
Db	202	CCTCTTAGATGCTTCTTATCTGCTCAAGTGCACTGTGGATGTGGAGAAGAAATGGC	261				
QY	1960	CTGTCCTCTATTAGCCAGGTGTTTTATTACCAAGATGTTTAAAGTCAGAGAAGAAATGTTAT	2019				
Db	262	TTCTCTCTAGTAAATCAGATCTATTACTACCAATAATGTGAAATGCGAGCGAGAGATGTTCT	321				
QY	2020	GATAAAGATCATCATGCTTCAGATGTGTCATCTTTAATGATCCCATTAAGTTCCTTG	2079				
Db	322	GACAAGCATAGTATGCTTCAGACAGGTGTTCTCCATGATGGATCCAAACCACTTCCCTG	381				
QY	2080	TTACTGGTACTTTCAGAGGTATGAACCTGCGAGGCTTTTAAACAAGACCATTATCTACAAA	2138				
Db	382	ATGATCATGCTCAGCGCTTTGAACCTCTATCAGCTCTTCAGCAGCGCTGACATATGGGAG	441				
QY	2139	-----AGACCAGATTTGATTAAACAATATAATACACTAATAGAA	2178				
Db	442	AGATTTCAGTTCTGAGGTTTACCCATAAGAGACGTCGTTTCAGCAGAACCAACACTCTGTATCGAA	501				
QY	2179	GAATCTCTTCAGTCTCATCTATATTGTGGGTGAGCGTTTACTACCTGAGTGGGAAAT	2238				
Db	502	GAGATGCTCTACTCTATCATCATGCTGTGTGGAGAAGATTCAAACCTGGGTGGACAG	561				
QY	2239	GTGACCAAAAGAGGTCACAATGAGAGAAATCATTTACTTGGTTTGCATTTGAACCCATG	2298				
Db	562	GTGGCTGCCACAGATCAAAATCAAGAGGGAGATTATCCATCAGCTGAGCATCAAACTATG	621				
QY	2299	CCACACAGTGCCTATTCGCAAAATTTACCTGAGATGAAAATATGAAACTGGCTTAGAG	2358				
Db	622	GCTCACAGCGAGCGTGTGGAAGTCTCTGCTGGAAGATGGAACAAGGAGACCGSCATGGAG	681				
QY	2359	AATGTCATAAAACAAAGTGCCACATTTAAGAACACAGGTGTATCAGGCGCATGGAGTTTAT	2418				
Db	682	AGCGTCATCGAGTCCTAGCACAATTCAGAAAACCTGGCCTCACAGCAGCAGGCATGTAT	741				
QY	2419	GAACATAAAGATGAATCACTGAAAGACTTCAATATATGTACTTTTATCATTTACTCCAAAAC	2478				
Db	742	GAGCTGAAGCCAGAGTGTGCCAAAGAGTTTCAACCTGTATTTTATCATTTCTCCAGGCA	801				



/tissue_type="melanotic melanoma"									
/lib_host="DH10B (phage-resistant)"									
/clone_lib="NIH_MGC_72"									
/note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI;									
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.									
Average insert size 2 kb. Library constructed by Life									
Technologies."									
BASE COUNT	261 a	146 c	204 g	250 t	4	others			
ORIGIN									
Query Match	14.8%	Score 771.4;	DB 13;	Length 865;					
Best local Similarity	99.08;	Pred. No. 1e-147;							
Matches 786;	Conservative 0;	Mismatches 7;	Indels 1;	Gaps 1;					
QY	1601	TCCAAGAGTGGTGGCTTGGATCAGAACTCTTACTTGGCTTATAAAGAAATGTCACA	1660						
Db	8	TCCGAGAGTGGTGGCTTGGATCAGAACTCTTACTTGGCTTATAAAGAAATGTCACA	67						
QY	1661	AAGCTGTGTAGAGTGCAGTACCAGTTTCATATCTAGTACCAAGACAGTAGTACAAATCGT	1720						
Db	68	AAGCTGTGTAGAGTGCAGTACCAGTTTCATATCTAGTACCAAGACAGTAGTACAAATCGT	127						
QY	1721	GTGCACATAGTTTGGAAACAAGTCCTACAGAGTATCTGAGGATCTTGTAAAGCATACATC	1780						
Db	128	GTGCACATAGTTTGGAAACAAGTCCTACAGAGTATCTGAGGATCTTGTAAAGCATACATC	187						
QY	1781	TGCCACTCTCTAGSACCCCTTGGCTGCTTCATCTGACGTTTAAAGCAGGCTGGGTGCTGTTT	1840						
Db	188	TGCCACTCTCTAGSACCCCTTGGCTGCTTCATCTGACGTTTAAAGCAGGCTGGGTGCTGTTT	247						
QY	1841	CAAGACTGCGATGAATTTGCTCTCTTTTGAGGACCTTCAAGTAGAGGTACTAGTGGAAATATC	1900						
Db	248	CAAGACTGCGATGAATTTGCTCTCTTTTGAGGACCTTCAAGTAGAGGTACTAGTGGAAATATC	307						
QY	1901	CTTTACGTTGCTGGGTGGTTGGTTGCCCAGTTGTTGCTGAGATGTGCCGAAGAAATGGAC	1960						
Db	308	CTTTACGTTGCTGGGTGGTTGGTTGCCCAGTTGTTGCTGAGATGTGCCGAAGAAATGGAC	367						
QY	1961	TGTCCTCTTATAGCCAGGTGTTTATTACCAAGATGTTAAGTCGAGAGAAGAATGTATG	2020						
Db	368	TGTCCTCTTATAGCCAGGTGTTTATTACCAAGATGTTAAGTCGAGAGAAGAATGTATG	427						
QY	2021	ATAAGATATCATCATGCTTCAGATGGTGCATCTTTTAAATGGATCCCAATAAAGTCTTGT	2080						
Db	428	ATAAGATATCATCATGCTTCAGATGGTGCATCTTTTAAATGGATCCCAATAAAGTCTTGT	487						
QY	2081	TACTGGTACTTCAGAGTATGAATTCGCCAGGCTTTTACACAGACCATATCTACAAAG	2140						
Db	488	TACTGGTACTTCAGAGTATGAATTCGCCAGGCTTTTAAACAGACCATATCTACAAAG	547						
QY	2141	ACCAGGATTTGATTAACAATAATAACACTAATAGAAGAAATGCTTCAGGTCCTCATCT	2200						
Db	548	ACCAGGATTTGATTAACAATAATAACACTAATAGAAGAAATGCTTCAGGTCCTCATCT	607						
QY	2201	ATATTGGGTGAGCGTTATGTACCTGGAGTGGGAAATGTGACCAAGAGAGGTACAA	2260						
Db	608	ATATTGGGTGAGCGTTATGTACCTGGAGTGGGAAATGTGACCAAGAGAGGTACAA	667						
QY	2261	TGAGAGAAATCATTCACTTCGCTTTGGATTGAACCCATGCCACACAGTCCCATTCGCCAAA	2320						
Db	668	TGAGAGAAATCATTCACTTCGCTTTGGATTGAACCCATGCCACACAGTCCCATTCGCCAAA	727						
QY	2321	ATTTTACCTGAGAATGAAAATAATGAAAT--GGCTTTAGAGAAATGTCATAAACAAAGTGCC	2379						
Db	728	ATTTTACCTGAGAATGAAAATAATGGAACCTGGCTTANAGAATGTCTATAAACAAAGTGCC	787						
QY	2380	ACATTTTAAGAAACC	2393						
Db	788	ACATTTTAAGAAAC	801						

RESULT 5  
BX104087

LOCUS	BX104087	712 bp	mRNA	linear	EST 06-FEB-2000
DEFINITION	BX104087 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:742663, mRNA sequence.				
ACCESSION	BX104087				
KEYWORDS	BX104087.1 GI:27846028				
SOURCE	EST.				
ORGANISM	Homo sapiens (human)				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	1 (bases 1 to 712) Ebert,L., Heil,O., Hennig,S., Neubert,P., Partsch,E., Peters,M., Radelof,U., Schneider,D. and Korn,B.				
TITLE	Human Unigeneset - RZPD3				
JOURNAL	Unpublished				
COMMENT	Contact: Ina Rolfs RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH Im Neuenheimer Feld 580, D-69120 Heidelberg, Germany RZPD: IMAGp998f081824. RZPDLIB: I.M.A.G.E. cDNA Clone Collection; Human Unigeneset - RZPD3 (RZPDLIB No.972) http://www.rzpd.de/CloneCards/cgi-bin/showlib.pl.cgi/response?libNo=972 Contact: Ina Rolfs RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH Heuberweg 6, D-14059 Berlin, Germany Tel: +49 30 32639 101 Fax: +49 30 32639 111 www.rzpd.de This clone is available royalty-free from RZPD; contact RZPD (clone@rzpd.de) for further information. Seq primer: M13r, Primer sequence: TTTTACACAGGAACACAGCTATGAC.				
FEATURES	Location/Qualifiers				
source	1..712				
	/organism="Homo sapiens"				
	/mol_type="mRNA"				
	/db_xref="taxon:9606"				
	/clone="IMAGp998f081824" ; IMAGE:742663"				
	/sex="male"				
	/lab_host="DH10B"				
	/clone_lib="Soares_testis_NHT"				
	/notes="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was prepared from mRNA obtained from Clontech Laboratories, Inc., and primed with a Not I - oligo(dT) primer [5', TGTTACATCTGAAGTGGAGCGGCCCAATTTTCTTTTCTTTT 3']. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization to Cot5, and was constructed by Bento Soares and M. Fatima Bonaldo."				
BASE COUNT	174 a	164 c	173 g	201 t	
ORIGIN					
Query Match	13.6%; Score 708.8; DB 13; Length 712;				
Best Local Similarity	99.7%; Pred. No. 7.1e-135;				
Matches	710; Conservative	0; Mismatches	2; Indels	0; Gaps	0;
QY	4433	CTGGCTGGATTTGTGGGCTCTCACTGAAGATGGCATCACCCCTTATCTTCGCTGTGCTG	4492		
Db	1	CTGGCTGGATTTGTGGGCTCTCACTGAAGATGGCATCACCCCTTATCTTCGCTGTGCTG	60		
QY	4493	CATTGTTTTTCCACTATTTACTTGGGGTAACTCGCGCTGAGGAACTCATACCAATTCGTG	4552		
Db	61	CATTGTTTTTCCACTATTTACTTGGGGTAACTCGCGCTGAGGAACTCATACCAATTCGTG	120		
QY	4553	CAGAAGGAGAGTACAGTGCACCTGTAGCTATCTATCTTTTACCTACAAATTTGTTCTGTC	4612		
Db	121	CAGAAGGAGAGTACAGTGCACCTGTAGCTATCTATCTTTTACCTGCAAAATTTGTTCTGTC	180		
QY	4613	TCTTCCAGGAATATTTGGGATCTGTGAAGCGCCCTTGCTCCAGAGCGGTGTGCAGATCCTG	4672		
Db	181	TCTTCCAGGAATATTTGGGATCTGTGAAGCGCCCTTGCTCCAGAGCGGTGTGTGCAGATCCTG	240		

```
QY 4673 CTTTACTAACTGTTTGAAGCAAAAAACACCGTGGTGCAGGTACCTAGAAAAAGAAATA 4732
Db 241 CTTTACTAACTGTTTGAAGCAAAAAACACCGTGGTGCAGGTACCTAGAAAAAGAAATA 300
QY 4733 GTTTCATAGAGTCTCTGATGACTATAGTGGCTCCTGAATCAAGTCTCTATTCAGGT 4792
Db 301 GTTTCATAGAGTCTCTGATGACTATAGTGGCTCCTGAATCAAGTCTCTATTCAGGT 360
QY 4793 GCCCAGCTGTCAGATGATGAGCAAAAGCATCTCTCTGCTGCTCTCTCTCTCTCTCTCT 4852
Db 361 GCCCAGCTGTCAGATGATGAGCAAAAGCATCTCTCTGCTGCTCTCTCTCTCTCTCTCT 420
QY 4853 TACTATGTTCTCAGAACATTTGCTGCCAGGAAATTTGTAACGGGGAAGGTTGGAGCTT 4912
Db 421 TACTATGTTCTCAGAACATTTGCTGCCAGGAAATTTGTAACGGGGAAGGTTGGAGCTT 480
QY 4913 GCATTTTTCAGCACTTCACTGTGGAGCGGAGTCTGCATTTTCTTAAATCAGAGAT 4972
Db 481 GCATTTTTCAGCACTTCACTGTGGAGCGGAGTCTGCATTTTCTTAAATCAGAGAT 540
QY 4973 GCCGAGTGGTCTGTTGAAAGTAAAGCCAGAGGCTGTCCTATCCAGCTCCTTACTTGG 5032
Db 541 GCCGAGTGGTCTGTTGAAAGTAAAGCCAGAGGCTGTCCTATCCAGCTCCTTACTTGG 600
QY 5033 ATGAATATGGAAACACACCTGCTGCTGAGAGGGCAACCCCTTCAATTTATCTCTG 5092
Db 601 ATGAATATGGAAACACACCTGCTGCTGAGAGGGCAACCCCTTCAATTTATCTCTG 660
QY 5093 AGCGGTATCGGAAGTCCATTTGTTGTCGCAACACACTGCATTTATAGAAGA 5144
Db 661 AGCGGTATCGGAAGTCCATTTGTTGTCGCAACACACTGCATTTATAGAAGA 712
```

## RESULT 6

## AI929033

## LOCUS

DEFINITION au64c10 y1 Schneider fetal brain 00004 Homo sapiens cDNA clone  
IMAGE:2519538 5' similar to TR:070481 070481 UBIQUITIN-PROTEIN  
LIGASE E3 COMPONENT N-RECOGNIN ;, mRNA sequence.

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

Homo sapiens (human)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 682)  
Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,  
Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin  
J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B.,  
White, Y., Wyllie, T., Waterston, R. and Wilson, R.

## TITLE

## JOURNAL

## COMMENT

Unpublished  
Other\_ESTS: au64c10.x1  
Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@wustl.wustl.edu

This clone is available royalty-free through LNL ; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Possible reversed clone: similarity on wrong strand

Seq primer: -40RP from Gibco

High quality sequence stop: 450.

## FEATURES

## source

Location/Qualifiers  
1..682  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2519538"  
/sex="male"  
/tissue\_type="frontal lobe"

/dev\_stage="5 months post-conception"  
/lab\_host="DH10B"  
/clone\_lib="Schneider fetal brain 00004"  
/note="Organ: brain; Vector: pBluescript SK (Stratagene);  
Site.1: SstI; Site.2: XhoI; Double-stranded cDNA was  
prepared from human fetal brain tissue. 5' and 3'  
adaptors were used in cloning as follows: 5' adaptor  
sequence:  
5'-GAGAGAGAGAGAGCTCAAGGATCCTTAATAATCCGCCGCCGCC-3'  
and 3' adaptor sequence:  
5'-GAGAGAGAGAGCTCGATTTTTTTTTTTTTTT-3'. The library was  
size-selected for >0.5 kb inserts and has an average  
insert size estimated at 1.2 kb. This library was  
constructed using the CAP-trapper method for full-length  
enrichment and has not undergone amplification. Library  
was constructed by Dr. Claudio Schneider (LNCIB-Area  
Science Park, Trieste, Italy)."  
BASE COUNT 178 a 155 c 166 g 182 t 1 others  
ORIGIN

Query Match 12.9%; Score 673; DB 9; Length 682;  
Best Local Similarity 99.1%; Pred. No. 1.6e-127;  
Matches 676; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
QY 4521 AACTCGCGCTGAGGAAGTGCATACCAATTCGCAGAGGAGAGTACAGTGCACCTCTGTAG 4580  
Db 1 AACTCGCGCTGAGGAAGTGCATACCAATTCGCAGAGGAGAGTACAGTGCACCTCTGTAG 60  
QY 4581 CTATCTATCTTTACCTACAAATTTGTTCTGCTCTTCCAGGAATATTTGGGATACTGTAA 4640  
Db 61 CTATCTATCTTTACCTACAAATTTGTTCTGCTCTTCCAGGAATATTTGGGATACTGTAA 120  
QY 4641 GCCCTTGTCTCCAGAGCGGTGTGCAGATCCTGCTTACTAACTGTTTGAAGCAAAAAA 4700  
Db 121 GCCCTTGTCTCCAGAGCGGTGTGCAGATCCTGCTTACTAACTGTTTGAAGCAAAAAA 180  
QY 4701 CACCGTGTGTCAGTACCCCTAGAAAAAGAAATATTTGATAGAGCTTCTCTGATGACTATAG 4760  
Db 181 CACCGTGTGTCAGTACCCCTAGAAAAAGAAATATTTGATAGAGCTTCTCTGATGACTATAG 240  
QY 4761 CTGCTCTCTGAATCAAGCTTCTCATTTTCAGGTGCCCGCTGTCAGATCATGCGCAAA 4820  
Db 241 CTGCTCTCTGAATCAAGCTTCTCATTTTCAGGTGCCCGCTGTCAGATCATGCGCAAA 300  
QY 4821 GCATCTGTCTCTGCTCTTCTTCTGTTGGGCTATATCTCTCAGAACATTTGCTGCCA 4880  
Db 301 GCATCTGTCTCTGCTCTTCTTCTGTTGGGCTATATCTCTCAGAACATTTGCTGCCA 360  
QY 4881 GGAATTTGTAACGGGGAAGAGGTTGGAGCTTGCATTTTTCACGCACTTCACTGTGGAGC 4940  
Db 361 GGAATTTGTAACGGGGAAGAGGTTGGAGCTTGCATTTTTCACGCACTTCACTGTGGAGC 420  
QY 4941 CGGAGTCTGCATTTTCTTAAATCAGAGATCCGAGTGGTCTGTTGAAGTAAAGC 5000  
Db 421 CGGAGTCTGCATTTTCTTAAATCAGAGATCCGAGTGGTCTGTTGAAGTAAAGC 480  
QY 5001 CAGAGCTGTGCTATCTCAGCTCCTTACTTGGATGAATATGGAGAACACAGACCTGGCCT 5060  
Db 481 CAGAGCTGTGCTATCTCAGCTCCTTACTTGGATGAATATGGAGAACACAGACCTGGCCT 540  
QY 5061 GAAGAGGGCAACCCCTTTCATTTATCTCTGAGCGGTATCGGAAGCTCCATTTGGTCTG 5120  
Db 541 GAAGAGGGCAACCCCTTTCATTTATCTCTGAGCGGTATCGGAAGCTCCATTTGGTCTG 600  
QY 5121 GCAACAACTGCATTTATAGAGAGATTCGTAGAGCCAGACATTAATCAGATGTTATT 5180  
Db 601 GCAACAACTGCATTTATAGAGAGATTCGTAGAGCCAGACATTAATCAGATGTTATT 660  
QY 5181 TGGATTCAACTGGCAGTTACTG 5202  
Db 661 TGGATTCAACTGGCAGTTACTG 682



```

RESULT 7
BU618516/c
LOCUS
DEFINITION
  UI-H-FH1-bfk-c-21-0-UI.s1 NCI_CGAP_FH1 Homo sapiens cDNA clone
  UI-H-FH1-bfk-c-21-0-UI 3', mRNA sequence.
ACCESSION
  BU618516
VERSION
  BU618516.1 GI:23284731
KEYWORDS
  EST.
SOURCE
  Homo sapiens (human)
ORGANISM
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  1 (bases 1 to 668)
REFERENCE
  NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
  Tumor Gene Index
  Unpublished
  Contact: Robert Strausberg, Ph.D.
  Email: cgaps-remail.nih.gov
  Tissue Procurement: James Martin
  cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
  DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
  Clone Distribution: Clone distribution information can be obtained
  from Dr. M. Bento Soares, bento-soares@uiowa.edu
  Seq primer: M13 FORWARD
  POLYA=Yes.
FEATURES
  source
    1..668
    /location=Qualifiers
    /organism="Homo sapiens"
    /mol_type="mRNA"
    /db_xref="taxon:9606"
    /clone="UI-H-FH1-bfk-c-21-0-UI"
    /tissue_type="Cell Line"
    /dev_stage="Adult"
    /lab_host="DH10B (Life Technologies)"
    /clone_lib="NCI_CGAP_FH1"
    /note="Organ: Chondrosarcoma; Vector: pT7T3-Pac (Pharmacia)
    with a modified polylinker; Site_1: EcoR I; Site_2: Not
    I; NCI_CGAP_FH1 is a normalized cDNA library obtained from
    a cell line derived from grade I chondrosarcoma tissue.
    The library was constructed and normalized according to
    Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
    1996. First strand cDNA synthesis was primed with an
    oligo-dT primer containing a Not I site. Double stranded
    cDNA was ligated to an EcoR I adaptor, digested with Not
    I, and cloned directionally into pT7T3-Pac vector. The
    oligonucleotide used to prime the synthesis of
    first-strand cDNA contains a library tag sequence that is
    located between the Not I site and the (dT)18 tail. The
    sequence tag for this library is AGAATCCGGC. The cell line
    was provided by Dr. James Martin from the University of
    Iowa."
    TAG_LIB=UI-H-FH1
    TAG_TISSUE=Human Chondrosarcoma Cell Line CS8 - Grade 1
    TAG_CHROM=Chondrosarcoma
    TAG_SEQ=AGAATCCGGC"
  BASE COUNT 180 a 143 c 110 g 235 t
  ORIGIN

Query Match 12.8%; Score 664.8; DB 13; Length 668;
Best Local Similarity 99.7%; Pred. No. 7.7e-126;
Matches 666; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1867 GAGGACTTCAAGTAGAGGTACTAGTGGATATCCTTACGTGTCGTGGTGGTGGCC 1926
      |||||||
Db 668 GAGGACTTCAAGTAGAGGTACTAGTGGATATCCTTACGTGTCGTGGTGGTGGCC 609
      |||||||
QY 1927 CAGGTTGTTGCTCAGATGTGGCGAAGAAATGGAGTCTCTATTAGCCAGGTGTTTAT 1986
      |||||||
Db 608 CAGGTTGTTGCTCAGATGTGGCGAAGAAATGGAGTCTCTATTAGCCAGGTGTTTAT 549
      |||||||
QY 1987 TACCAAGATGTTAAGTGCAGAGAGAAATGTATGATAAAGATATCATCATGCTTCAGATT 2046
      |||||||

```

```

Db 548 TACCAAGATGTTAAGTGCAGAGAGAAATGTATGATAAAGATATCATCATGCTTCAGATT 489
      |||||||
QY 2047 GGTGCATCTTTAATGGATCCCAATAAGTCTCTTGTACTGGTACTTCCAGAGGTATGAACCT 2106
      |||||||
Db 488 GGTGCATCTTTAATGGATCCCAATAAGTCTCTTGTACTGGTACTTCCAGAGGTATGAACCT 429
      |||||||
QY 2107 GCCGAGGCTTTTAAACAAGACCATACTACAAAAGACCAGGATTTGATTAACAATAATAAT 2166
      |||||||
Db 428 GCCGAGGCTTTTAAACAAGACCATACTACAAAAGACCAGGATTTGATTAACAATAATAAT 369
      |||||||
QY 2167 ACACATAATAGAGAAATGCTTCAGGTCCTCATCTATATTGTTGGTGAGCGTTATGTACCT 2226
      |||||||
Db 368 ACACATAATAGAGAAATGCTTCAGGTCCTCATCTATATTGTTGGTGAGCGTTATGTACCT 309
      |||||||
QY 2227 GGAGTGGGAAATGTGACCAAAAGAGGTCACAATGAGAGAAATCATTCACCTGCTTTTGC 2286
      |||||||
Db 308 GGAGTGGGAAATGTGACCAAAAGAGGTCACAATGAGAGAAATCATTCACCTGCTTTTGC 249
      |||||||
QY 2287 ATTGAACCCATGCCACACAGTGCCTATGCAAAAATTTTACCTGAGAAATGAAAATAATGAA 2346
      |||||||
Db 248 ATTGAACCCATGCCACACAGTGCCTATGCAAAAATTTTACCTGAGAAATGAAAATAATGAA 189
      |||||||
QY 2347 ACTGGCTTAGAGATGTCTATAAACAAGTGGCCACATTTAAGAAACACAGGTATATCAGGC 2406
      |||||||
Db 188 ACTGGCTTAGAGATGTCTATAAACAAGTGGCCACATTTAAGAAACACAGGTATATCAGGC 129
      |||||||
QY 2407 CATGGAGTTTATGAACATAAAGATGAATCAGTCAAGACTTCAATATGTACTTTTATCAT 2466
      |||||||
Db 128 CATGGAGTTTATGAACATAAAGATGAATCAGTCAAGACTTCAATATGTACTTTTATCAT 69
      |||||||
QY 2467 TACTCCAAAACCCAGCATAGCAAGGCTGAACATATGCAAGAGAAAGAGAAACAAAGAA 2526
      |||||||
Db 68 TACTCCAAAACCCAGCATAGCAAGGCTGAACATATGCAAGAGAAAGAGAAACAAAGAA 9
      |||||||
QY 2527 AACAAAGA 2534
      |||||||
Db 8 AACAAAAA 1

RESULT 8
BU6534574
LOCUS
DEFINITION
  602553425F1 NIH_MGC_77 Homo sapiens cDNA clone IMAGE:4663182 5',
  mRNA sequence.
ACCESSION
  BU6534574
VERSION
  BU6534574.1 GI:13526116
KEYWORDS
  EST.
SOURCE
  Homo sapiens (human)
ORGANISM
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  1 (bases 1 to 756)
REFERENCE
  NIH-MGC http://mgc.nci.nih.gov/.
  National Institutes of Health, Mammalian Gene Collection (MGC)
  Unpublished
  Contact: Robert Strausberg, Ph.D.
  Email: cgaps-remail.nih.gov
  Tissue Procurement: CLONTECH Laboratories, Inc.
  cDNA Library Preparation: CLONTECH Laboratories, Inc.
  cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
  DNA Sequencing by: Incyte Genomics, Inc.
  Clone Distribution: MGC clone distribution information can be
  found through the I.M.A.G.E. Consortium/LLNL at:
  http://image.llnl.gov
  Plate: LNCMI465 row: a column: 07
  High quality sequence stop: 751.
FEATURES
  source
    1..756
    /organism="Homo sapiens"
    /mol_type="mRNA"
    /db_xref="taxon:9606"
    /clone="IMAGE:4663182"

```

/lab\_host="DH10B (T1 phage-resistant)"  
/clone\_lib="NIH\_MGC\_77"  
/note="Organ: lung; Vector: pDNR-LIB (Clontech); Site\_1:  
SfiI (ggcccctggcc); Site\_2: SfiI (ggccattggcc); 5' and  
3' adaptors were used in cloning as follows: 5' adaptor  
sequence: 5'-CACGCCATTATGCC-3' and 3' adaptor sequence:  
5'-ATTCTAGAGCGCGGCGGCACATG-DT(30)BN-3' (where B = A,  
C, or G and N = A, C, G, or T). Average insert size 1.9  
kb (range 0.5-4.0 kb). 12/15 colonies contained inserts  
by PCR. This library was enriched for full-length clones  
and was constructed by Clontech Laboratories (Palo Alto,  
CA). Note: this is a NIH\_MGC Library."  
BASE COUNT 274 a 138 c 152 g 192 t  
ORIGIN

Query Match 12.6%; Score 656.6; DB 10; Length 756;  
Best Local Similarity 95.9%; Pred. No. 3.8e-124;  
Matches 730; Conservative 0; Mismatches 19; Indels 12; Gaps 5;  
QY 2096 GGTATGAACCTGGCGAGCTTTTACAAAGACCATATCTACAAAGACCAAGATTTGATTA 2155  
Db 1 GGTATGAACCTGGCGAGCTTTTACAAAGACCATATCTACAAAGACCAAGATTTGATTA 60  
QY 2156 AACAAATATACACTAATAGAAGAAATGCTTCAGGTCTCATCTATATTTGGGTGAGC 2215  
Db 61 AACAAATATACACTAATAGAAGAAATGCTTCAGGTCTCATCTATATTTGGGTGAGC 120  
QY 2216 GTTATGTACCTGGAGTGGGAATGTGACCAAGAGAGGTGACAAATGAGAAATCATTC 2275  
Db 121 GTTATGTACCTGGAGTGGGAATGTGACCAAGAGAGGTGACAAATGAGAAATCATTC 180  
QY 2276 ACTTGTCTTGGATTGAACCCATGCCACAGTGCATTCGCAAAATTTACCTGAGAATG 2335  
Db 181 ACTTGTCTTGGATTGAACCCATGCCACAGTGCATTCGCAAAATTTACCTGAGAATG 240  
QY 2336 AAAATAATGAACCTGGCTTAGAGATGTCATAAACAAGTGGCCACATTTAAGAACACAG 2395  
Db 241 AAAATAATGAACCTGGCTTAGAGATGTCATAAACAAGTGGCCACATTTAAGAACACAG 300  
QY 2396 GTGTATCAGGCGCATGGAGTTTATGAACCTAAAGATGAATCACTGAAAGACTTCAATATGT 2455  
Db 301 GTGTATCAGGCGCATGGAGTTTATGAACCTAAAGATGAATCACTGAAAGACTTCAATATGT 360  
QY 2456 ACTTTTATCATTTACTCCAAACCCAGATAGCAAGGCTGACATATGACAGAAAGAA 2515  
Db 361 ACTTTTATCATTTACTCCAAACCCAGATAGCAAGGCTGACATATGACAGAAAGAA 420  
QY 2516 GAAACAAAGAAACAAAGATGAAG-----CATTCGCCGCCACCCACCTCCTGAAATCT 2569  
Db 421 GAAACAAAGAAACAAAGATGAAGTAAACATTTGCCGCCACCCACCTCCTGAAATCT 480  
QY 2570 GCCCTGCTTTCAGCAAACTGATTAACTTCTCAACTGTGTATAT-CATGATGTACATTC 2628  
Db 481 GCCCTGCTTTCAGCAAACTGATTAACTTCTCAACTGTGTATATCCATGATGTACATTC 540  
QY 2629 AGGACCGTATTGACGGCGCAATAGACACAGATTTACTTGTGGACCGAGGGATGCTC 2688  
Db 541 AGGACCGTATTGACGGCGCAATAGACACAGATTTACTTGTGGACCGAGGGATGCTC 600  
QY 2689 CAATGGCTTTTCATATCTGCGATTTGGTGTCTTACTAGAGAGAACACAGCTTCAAAA 2748  
Db 601 CAATGGCTTTTCATATTTCTGGCATTTGGTGTCTTACTAGAGAGAACACAGCTTCAAAA 659  
QY 2749 GCTCCTGAAGAAGAGTAAACATTTGACCTTTTATCATAAGGCTTCAAGATTTGGGAAGTTCA 2808  
Db 660 GCTCCTGAAGAAGAGTAAACATTTGACCTTTTATCATAAGGCTTCAAGATTTGGGAAGTTCA 719  
QY 2809 GCCATGAATATACAAATGCTTTTGGAAAACTTCAAAAGAA 2849  
Db 720 G-CATGAATATACAA--TGTTTTGGAAAAATCAAAAGATT 756

RESULT 9

AK083320  
LOCUS  
DEFINITION  
Mus musculus 2 days neonate thymus cells cDNA, RIKEN  
full-length enriched library, clone: C92004H05 product: similar to  
(A PROTEIN SIMILAR TO UBIQUITIN-PROTEIN LIGASE E3 COMPONENT  
N-RECOGNIN (UBIQUITIN-PROTEIN LIGASE E3-ALPHA)) (FRAGMENT) [Homo  
sapiens], full insert sequence.  
AK083320  
VERSION AK083320.1 GI:26350448  
KEYWORDS HTC; CAP trapper.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1  
Carninci, P. and Hayashizaki, Y.  
High-efficiency full-length cDNA cloning  
Meth. Enzymol. 303, 19-44 (1999)  
99279253  
10349636  
2  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new genes  
Genome Res. 10 (10), 1617-1630 (2000)  
20499374  
11042159  
3  
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,  
Konno, H., Akiyama, J., Nishi, K., Katsunari, T., Tashiro, H., Itoh, A.,  
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,  
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,  
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohata, E., Watanabe, M.,  
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J.,  
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer  
Genome Res. 10 (11), 1757-1771 (2000)  
20530913  
11076861  
4  
Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,  
Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S.,  
Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I.,  
Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R.,  
Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T.,  
Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H.,  
Kuehl, P., Lewis, S., Matsuo, Y., Nikaide, I., Pesole, G., Tomita, M.,  
Quackenbush, J., Schriml, L. M., Staubli, F., Suzuki, R., Tomita, M.,  
Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H.,  
Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N.,  
Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bult, C.,  
Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D.,  
Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P.,  
Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P.,  
Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,  
Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H.,  
Toyo-oka, K., Wang, K. H., Weitz, C., Whittaker, C., Wilming, L.,  
Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawai, H., Kohtsuki, S.  
and Hayashizaki, Y.  
Functional annotation of a full-length mouse cDNA collection  
Nature 409 (6821), 685-690 (2001)  
21085660  
11217851  
5  
The FANTOM Consortium and the RIKEN Genome Exploration Research  
Group Phase I & II Team.  
Analysis of the mouse transcriptome based on functional annotation  
of 60,770 full-length cDNAs  
Nature 420, 563-573 (2002)  
6 (bases 1 to 3627)  
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,

AK083320  
LOCUS  
DEFINITION

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS

TITLE  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS

TITLE  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS

TITLE  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS

TITLE  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS



Qy	1736	AAACAAGTCCTACAGAGTATCTGAGGATCTTTGTAAGCATACATCTGCCACTCTCTAGGA	1795
Db	2242	AAACCATCAGATACTGTGTTTCCCAAGAAAAGTTTAGCATTTACCTCCCAATTTCTCGCT	2301
Qy	1796	CCCTTGCTGCTCTTCATGTACGTTTAAAGCAGGCTGGTGCTGTTTCAAGACATCCATGAAT	1855
Db	2302	TGCTTGCGAGTTTGCATGTATGTTTAAAGCAAAAGTGAAGTGGCATATAATTTCCAGAGC	2361
Qy	1856	TTGTGTCTTTTGGAGGACTTTCAAGTAGAGTACTAGTGGAAATATCCTTTACGTGTGTCTGG	1915
Db	2362	TCCTACCTCTAAGTGAACATGAGCCACCCTATGTTGATAGACATCTCTTAGATGTCTTG	2421
Qy	1916	TGTTGGTTGCCAGGTTTGTCTGAGATGTGGCGAAGAAATGGACGTGCTCTCTTTATTAGCC	1975
Db	2422	TCTTATGTGCTCAAGTGCATGCTGGGATGTGGAGAAGAAATGGCTTCTCTAGTAAATC	2481
Qy	1976	AGTGTTTTTATTACCAAGATGTTAACTGCAGAGAAGAAATGTATATAAGATATCATCA	2035
Db	2482	AGATCTATTACTACCATATGTGAATGCAAGCGAGAGATGTTGCACAAGGACATAGTGA	2541
Qy	2036	TGCTTCAGATTGGTGACTCTTTAATGGATCCCAATAAGTTTCTTGTACTGGTACTTCAGA	2095
Db	2542	TGCTTCAGACAGGTGCTCCATCATGAGCCCAACCACTTCTCTGATGATCATGCTCAGCC	2601
Qy	2096	GGTATGAACATTGGCGAGGCTTTTAAACAAGACCATATCTACAAA-----	2138
Db	2602	GCTTTGAACCTCTATCAGCTCTTCAGCACGCGCTGACTATGGGAAGAGATTCAGTCTCGAGG	2661
Qy	2139	----AGACCAGGATTTGATTAAACAATATATACTAATAGAGAAGAAATGCTTCAGGTCC	2194
Db	2662	TTACCCNATAGGACGCTTTCACACAGAACACACTCTGATCGAAGAGATGCTCTACCTCA	2721
Qy	2195	TCATCTATATTGTGGGTGAGCGTTATGTACTCGAGTGGGAAATGTGACCAAAAGAGAGS	2254
Db	2722	TCATCATGCTTGTGGGAGAAAGATTCAACCTGGGTTGGACAGGTGGCTGCCACAGATG	2781
Qy	2255	TCCAATGAGAGAAATCATTTCACTTGTTCATTGAAACCCATGCCACACAGTGCACATTG	2314
Db	2782	AAATCAAGGGGAGATTATCCATCAGTTGAGCATCAAGCCATGGGCTCACAGTGAAGCTGG	2841
Qy	2315	CCAAAAATTTACCTGAGAATGAAATTAATGAAACTGGCTTAGAGAATGTCTAAACAAAG	2374
Db	2842	TGAAGTCTCTGCCTGAAGATGAGAACAGGAGACCGCATGGAGAGCGTCATCGAGTCCG	2901
Qy	2375	TGGCCACATTTAGAAAACAGGTGTATCAGGCCATGGAGTTTATGAACATAAAGATGAAT	2434
Db	2902	TTGCACATTTCAAGAAACCTTGGGCTCACAGGGCGAGGCATGTATGAGCTGAAGCCAGAGT	2961
Qy	2435	CACGTGAAGACTTCAATATGCTATTTATCATTTACTTCCAAAACCCAGCATAGCAAGGCTG	2494
Db	2962	GTGCCAAGAGTTCAACCTGATTTTATCATTTCTCCAGGGCAGAGCATGCCAAGGCAG	3021
Qy	2495	AACATATGCAGAGAAAGGAGAAAACAAGAAACAAAGATGAAGCATTTGCCGCCACCAC	2554
Db	3022	AGGAAGCTCAGCGGAAATTTGAAAGAGAAAATAAAGAAAGATACAGCACCTCCCTCCCGG	3081
Qy	2555	CACCTCCTGAATTTCTGCCCTGCTTTTCAGAAAAGTGATTAACCTTCTCAACTGTGATATCA	2614
Db	3082	CTTTTGCCACCGTTCTGCGCTTTGTTGCGAGTCTGGTTAAATCTTTCGAGTGTGAGGTCA	3141
Qy	2615	TGATGTACATTCACAGACCGTATTTGAGCGGCAATAGACACAGATTTCAACTTTGGA	2674
Db	3142	TGCTGTACATCATGGGAACGATCCTGCAGTGGGCTGTAGAGCATCAGGGGTCTGCGCTGGT	3201
Qy	2675	CCGAAGGGATGCTCCAAATGGCTTTTCATATTTCTGGCATTTGGGTTTACTAGAGAAGAAC	2734
Db	3202	CAGAGTCCATGCTACAGAGGCTGCTGATTTGATCGGATGGCTCTCCAGGAAGAGAGC	3261
Qy	2735	AACAGCTTCAAAAGCTCTCGAAGAGAAGT---AACATTTGACTTTTATCATAGGCTT	2791
Db	3262	ACCACTTGGAGAGCGCGTGGAGGGCACCTGGACAGCTTTCACGTCTCACACAGAAGATTT	3321

Qy	2792	CAGATGTGGGAAGTTCA	-----GCCATGAATATACAAATGCTTTTGGAAAAACTCA	2842
Db	3322	CAAAAGCCTGTGTATGGACCA	CAATAAATCCCGAGACCTCTAGCTATGCTGGAGACCTTGC	3381
Qy	2843	AAGGAATTCCTCCAGTTAGA	AGCCAGAACGACATGATAACGTGATACTTTCAGATGTTTG	2902
Db	3382	AGAACGCCCTCCCTGGAAG	CCCAAGGACATGATCAGGTGTTGCTAAAGATGTTTA	3441
Qy	2903	ACACAGTGAAGCGCATTAAG	AGAAAAATCTTTGTTTAATTGTAGCAACCACATCAGGATCGG	2962
Db	3442	ATGCAATTAAGAGATAAGAG	-----GTGTTCAATCCAGCAGCCCTGTGTGCCGAGGCGG	3495
Qy	2963	AATCTATTAGAATGATGAGAT	TACTCATGATTAAGAAAAAGCAGAACGAAAAAGAAAAG	3022
Db	3496	AGGGAACCAATAATGGAGAG	GAGGTCAAGAGACAAGGACAAAGCAGGAGAGAAAAAGAAAAG	3555
Qy	3023	CTGAAGCTGTAGGCTACATG	CGCAGAAAGATCATGGCTCAGATGTCTGCTTACAGAAA	3082
Db	3556	CCGAGATCGCAGACTCGCC	GGGAGGAAGATCATGGCCAGATGTTTGAGATCGACGGCG	3615
Qy	3083	ACTTCATTGA	3092	
Db	3616	ACTTCATTGA	3625	
RESULT 10				
LOCUS	60279907Af1	NCI_CGAP_Mam4	Mus musculus	cdna linear EST 29-MAY-2001
DEFINITION	60279907Af1	NCI_CGAP_Mam4	Mus musculus	cdna linear EST 29-MAY-2001
ACCESSION	60279907Af1	NCI_CGAP_Mam4	Mus musculus	cdna linear EST 29-MAY-2001
VERSION	60279907Af1	NCI_CGAP_Mam4	Mus musculus	cdna linear EST 29-MAY-2001
KEYWORDS	60279907Af1	NCI_CGAP_Mam4	Mus musculus	cdna linear EST 29-MAY-2001
SOURCE	60279907Af1	NCI_CGAP_Mam4	Mus musculus	cdna linear EST 29-MAY-2001
ORGANISM	60279907Af1	NCI_CGAP_Mam4	Mus musculus	cdna linear EST 29-MAY-2001
REFERENCE	60279907Af1	NCI_CGAP_Mam4	Mus musculus	cdna linear EST 29-MAY-2001
AUTHORS	60279907Af1	NCI_CGAP_Mam4	Mus musculus	cdna linear EST 29-MAY-2001
TITLE	60279907Af1	NCI_CGAP_Mam4	Mus musculus	cdna linear EST 29-MAY-2001
JOURNAL	60279907Af1	NCI_CGAP_Mam4	Mus musculus	cdna linear EST 29-MAY-2001
COMMENT	60279907Af1	NCI_CGAP_Mam4	Mus musculus	cdna linear EST 29-MAY-2001
FEATURES	60279907Af1	NCI_CGAP_Mam4	Mus musculus	cdna linear EST 29-MAY-2001
source	60279907Af1	NCI_CGAP_Mam4	Mus musculus	cdna linear EST 29-MAY-2001
BASE COUNT	194	a	177	c 207 g 219 t
ORIGIN	194	a	177	c 207 g 219 t

CONTACT: Robert Strausberg, Ph.D.  
Email: c9apbs-i@mail.nih.gov  
Tissue Procurement: NCI  
CDNA Library Preparation: Michael Brownstein Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LCM3047 row: f column: 07  
High quality sequence stop: 555.

high quarry sequence seq: 555.

FEATURES

Location/Qualifiers
1. .774

Query Match 11.5%; Score 596; DB 13; Length 774;  
Best Local Similarity 94.2%; Pred. No. 1.le-11;  
Matches 637; Conservative 5; Mismatches 28; Indels 6; Caps 2;

Qy 671 HSDHGTATATACAGCCTACAAAGAGCTCTTGACTGTGAGCTGCAGAGGCCCACTTGC 730  
.....|

QY	8531	ATGTAGAAAGTATTACACTCAGAGATTATGGCTCATCAGAAATTTGCTTTGGCGTCGGTT	910
D6	210	ATGCTTACAGACATAATATACATCTCACACAGATATATGCGCTACATCACAGAAATTTGCTTTGGCGTCGGTT	278

219	ATGGAGATGATATACACATCAGAGGATATATAGCCCTCATCAGGAAATTTACCTTACCGGCTCTTGGT	270
220		
221	CCCTGGATGACAAAAATATGAGCTATTCAAGTCACCTTTAGCCAGATCTTTTGGCAAGCAT	320
222		

Db 279 CCTGGATGAACAAATATATGAGCTATTCAAGTGACTTTAGGCAGATCTTTGGCAGCAT 338

Qy 971 GCCTTAGAGAAGAACTGACTCGGAGAAATCCCTGTCTCATAAGCAGGTTAATGCTTTGGG 1030

Db  
339 GCCTTAGAGAAGAACCTGACTCGGAGAATCCCTGTCTCATAGCAGGTTAATGCTTTGGG 398

Qy 1031 ATGCAAAGCTTTATAAAGGTGCCCGTAAGATCCTTCATGAATGATCTTCAGCAGTTTT 1090

Db 399 ATGCAAGCTTTATAAAGTGCCCGTAAGATCCTTCAATTCATCTTCAGCAGTTTT 458

QY 1091 TTATGGAGATGGAATACAAAAAACTCTTTGCTATGGAATTTGTGAAGTATTATAACAAC 1150

```

|||||
459 TTATGGAGTGAATACAAAACCTTTGCTATGGATTGCTGAAGTATTATTAACAC 518
QY TGCAGAAAGATATATATCAGTGATCATGACAGAGATATCTCTATTAACCTCAG 1210
|||||
519 TGCAGAAAGATATATATCAGTGATCATGACAGAGATATCTCTATTAACCTCAG 578
QY TTCAGATGTTTACTGTTCTTCTACTCTGGTCGACATCTTATGAAGACGAGATGTTATCT 1270
|||||
579 TTCAGATGTTTACTGTTCTTCTACTCTGGTCGACATCTTATGAAGACGAGATGTTATCT 635
|||||
1271 CTGTCATTACTGAACCTCTGCTAGAA--GTTTTACCTGAGTACTTGGACAGGACAAATA 1327
|||||
636 CTGTCATTACTGGAACCTCTGCTAGAAAGTTTACCCTGAGTACTTGGACCGGAACAAT 695
QY 1328 AATTCACCTTCAGGG 1343
Db 696 ACAAATTTCACTGG 711

```

RESULT 12  
 BY756516  
 LOCUS  
 DEFINITION BY756516 695 bp mRNA linear EST 17-DEC-2002  
 Clone IL0020G08 3', mRNA sequence.  
 ACCESSION BY756516.1 GI:27189572  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM

Mus musculus (house mouse)  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 695)  
 Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,  
 Nikaido, I., Osato, N., Saito, R., Suzuki, H., Yamanaka, I., Kiyosawa, H.,  
 Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C.,  
 Gojobori, T., Baldarelli, R., Hill, D.P., Bult, C., Hume, D.A.,  
 Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H., Batalov, S.,  
 Beisel, K.W., Blake, J.A., Brad, D., Brusic, V., Chothia, C., Corbani,  
 L.E., Cousins, S., Dalla, E., Dragani, T.A., Fletcher, C.F., Forrest,  
 A., Frazer, K.S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A.,  
 Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I.J.,  
 Jarvis, E.D., Kanai, A., Kawaj, H., Kawasawa, Y., Kedzierski, R.M.,  
 King, B.L., Konagaya, A., Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons,  
 P.A., Maglott, D.R., Maltais, L., Marchionni, L., McKenzie, L., Miki,  
 H., Nagashima, T., Numata, K., Okido, T., Pavan, W.J., Pertea, G.,  
 Pesole, G., Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D.,  
 Ramachandran, S., Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring,  
 B.Z., Ringwald, M., Sandelin, A., Schneider, C., Sempke, C.A., Setou,  
 M., Shinada, K., Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale,  
 R.D., Tomita, M., Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y.,  
 Watanabe, Y., Wells, C., Wilming, L.G., Wynshaw-Boris, A., Yanagisawa,  
 M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A.,  
 Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura,  
 M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K.,  
 Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii,  
 Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata,  
 K., Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander,  
 E.S., Rogers, J., Birney, E. and Hayashizaki, Y.

Analysis of the mouse transcriptome based on functional annotation  
 of 60,770 full-length cDNAs  
 Nature 420, 563-573 (2002)

22354683  
 12466851

CONTACT: Yoshihide Hayashizaki  
 Laboratory for Genome Exploration Research Group, RIKEN Genomic  
 Sciences Center (GSC), Yokohama Institute  
 The Institute of Physical and Chemical Research (RIKEN)  
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
 Tel: 81-45-503-9222  
 Fax: 81-45-503-9216  
 Email: genome-res@gsr.riken.go.jp/  
 URL: http://genome.gsc.riken.go.jp/

TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 COMMENT

# FEATURES

## source

Location/Qualifiers  
 1. 695  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="IL0020G08"  
 /dev\_stage="blastocyst"  
 /clone\_lib="RIKEN full-length enriched, blastocyst"  
 218 a 128 c 157 g 191 t 1 others

# BASE COUNT

## ORIGIN

Query Match 11.2%; Score 585.2; DB 14; Length 695;  
 Best Local Similarity 90.1%; Pred. No. 1.7e-109;  
 Matches 626; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

```

QY 1809 TCATGTAGCTTTAAGCAGGCTGGTCTCTTTTCAAGCTGATGATTTGTCCTTTTGA 1868
Db 1 TCATGTAGCTTTAAGCAGGCTGGTCTCTTTTCAAGCTGATGATTTGTCCTTTTGA 60
QY 1869 GCACCTTCAAGTAGAGGTACTAGTGAATATCTTTACCTTGTCTGTGTGTGTCGCCA 1928
Db 61 CGGCTTTCAAGTAGAGGTCTCTGGTGAGTACCGCTGCTGCTGCTGCTGCTGCTCA 120
QY 1929 GCTTCTGCTGAGATGTGGCGAAGAAATGGACTGTCTCTTATTAGCAGGTGTTTATTATTA 1988
Db 121 GCTTCTGCTGAGATGTGGCGAAGAAATGGACTGTCTCTTATTAGCAGGTGTTTATTATTA 180
QY 1989 CCAAGATGTTAAGTCAGAGAAATGATGATGAAGATATCATCATCATCATCATCATCATCAT 2048
Db 181 TCAAGATGTTAAGTCAGAGAAATGATGATGAAGATATCATCATCATCATCATCATCATCAT 240
QY 2049 TGCATCTTTAATGGATCCCAATGAATTTCTTTACTGGTACTTCAGAGTATGAACCTGC 2108
Db 241 AGCATCTAATATGGATCCCAATGAATTTCTTTACTGGTACTTCAGAGTATGAACCTGC 300
QY 2109 CGAGGCTTTTAAACAGACCATATCTCAAAAGACAGGATTTTCATTAAACAATATATAC 2168
Db 301 TGATGCTTTTAAACAGACCATATCTCAAAAGACAGGATTTTCATTAAACAGTATATAC 360
QY 2169 ACTAATAGAAGAAATGCTTCAGGCTCTCATCTATATTTGTGGGTGAGCGGTATGACCTGG 2228
Db 361 ATTATAGAAGAAATGCTTCAGGCTCTCATCTATATTTGTGGGTGAGCGGTATGACCTGG 420
QY 2229 AGTGGGAATGTGACCAAGAGAGGTCAATGAGAGAAATCATTCATCTGCTTTGTCAT 2288
Db 421 AGTGGGAATGTGACCAAGAGAGGTCAATGAGAGAAATCATTCATCTGCTTTGTCAT 480

```

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hashizume, W., Hayashida, K., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Koya, S., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Saito, R., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Takeda, Y., Waki, K., Watanabe, A., Muramatsu, M. and Hayashizaki, Y. Direct Submission

Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)  
 Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)  
 RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
 cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.  
 Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

QY 2289 TGAACCCATGCCACAGCTGCCATTGCCAAAATTTACCTGAGATGAATAATGAAC 2348  
DB 481 TGACCCCATGCCATAGTCCATGCCAGAACCTTACCTGAGAACGAAATAATGAAC 540  
QY 2349 TGGCTTAGAGATGTCATAAACAAGTGGCCACATTTAAGAAACCAGGTGTATCAGGCCA 2408  
DB 541 TGGCTTAGAGATGTCATAAACAAGTGGCCACATTTAAGAAACCAGGTGTATCAGGCCA 600  
QY 2409 TGGAGTTTGAACATAAAGATGAATCACTGAAGACATTCATATGTAATTTATCAATTA 2468  
DB 601 TGGAGTTTATGAATTTGAAGATGAATCACTGAAGACATTCATATGTAATTTATCAATTA 660  
QY 2469 CTCAAAACCCAGCATACAGGCTGAACATATGC 2503  
DB 661 TTCTAAACACAGCATACAGGCTGAACATATGC 695

RESULT 13  
BM114217/c

LOCUS  
DEFINITION L0801D04-3 NIA Mouse Newborn Brain cDNA Library Mus musculus cDNA  
clone L0801D04 3', mRNA sequence.

ACCESSION  
VERSION BM114217  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)

ORGANISM  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 703)  
Piao, Y., Kargul, G.J., Dudekula, D.B., Qian, Y., Pantano, S., Lim, M.K.  
and K.M.S.H.

TITLE  
JOURNAL Systematic Analyses of NIA Mouse Newborn Brain cDNA Library  
COMMENT Unpublished  
Contact: Dawood B. Dudekula  
Laboratory of Genetics  
National Institute on Aging/National Institutes of Health  
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA  
Email: cdna@igsun.grc.nia.nih.gov  
Plate: L0801 row: D column: 04  
Seq primer: -21M13 Forward  
High quality sequence stop: 703  
POLYA=Yes.

FEATURES  
source  
1..703  
Location/Qualifiers  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="nia:EST:L0801D04-3"  
/db\_xref="taxon:10090"  
/clone="L0801D04"  
/tissue\_type="Newborn Brain"  
/dev\_stage="Newborn"  
/lab\_host="DHL0B"  
/clone\_lib="NIA Mouse Newborn Brain cDNA Library"  
/note="Organ: brain; Vector: pSPORT1 (Invitrogen); Site: 1:  
Salt; Site: 2: NotI; Mouse cDNA project by the Laboratory  
of Genetics, National Institute on Aging (NIA), Intramural  
Research Program, NIH (http://igsun.grc.nia.nih.gov/cDNA).  
Double-stranded cDNAs were synthesized with an Oligo(dT)  
primer [Invitrogen: 5'-  
pGACTAGTCTAGATCGGAGCGCGCCCTTTT-3'] from 48  
microgram of total RNA, treated with T4 DNA polymerase,  
and purified by ethanol-precipitation. The cDNAs were  
ligated to Lene-linker LL-Sal3 (Ref. Development  
127:1737-1749 (2000) [PMID:10725249]), purified by  
phenol/chloroform, and separated from free linkers by  
Centricon 100. Then, the cDNAs were digested with SalI and  
NotI enzymes, and cloned into SalI/NotI site of pSPORT1  
plasmid vector. The DHL0B E. Coli host was transformed  
with ligation mixture by the standard chemical method. The  
average insert size is about 1.9 kb. The library was

BASE COUNT 180 a 158 c 128 g 237 t  
ORIGIN  
constructed by Yulan Piao (NIA)."  
Query Match 11.2%; Score 583.6; DB 12; Length 703;  
Best Local Similarity 90.1%; Pred. No. 3.7e-109;  
Matches 625; Conservative 0; Mismatches 69; Indels 0; Gaps 0;  
QY 1845 ACTGCATGAATTTGTCTCTTTTGAGGACTTTTCAAGTAGAGGTACTAGTGAATATCCTTT 1904  
DB 703 ACTGCATGAATTTGTCTCTTTTGAGGACTTTTCAAGTAGAGGTCTCTGTGGAGTACCCCT 644  
QY 1905 AGTTCCTCTGGTGTCTGTTGCCAGGTGTGTCTGAGATGTGGCGAAGAAATGACATGTC 1964  
DB 643 GCGCTGCCCTGGTCTGCTGCTCAGGTGTGTCTGAGATGTGGCGAAGAAACGGGCTCTC 584  
QY 1965 TCTTATTAGCCAGGTGTTTATTACCAAGATGTTTAAGTCAGAGAAGAAATGTATGATAA 2024  
DB 583 ACTCATCAGCCAGGTTTCTATTATCAAGATGTTTAATGCAGGAGGAATGTACGATAA 524  
QY 2025 AGATATCATCATGCTTTCAGATTTGGTGCATCTTTTAATGGATGCCAATTAAGTCTTGTACT 2084  
DB 523 AGATATCATCATGCTTTCAGATTTGGATGATCTATAATGGATGCCAACAAGTCTTGTACT 464  
QY 2085 GGTACTTTCAGAGGTATGAACCTTCCCGAGGCTTTTAAACAAGACCATATCTACAAAGACCA 2144  
DB 463 GGTACTTTCAGAGGTATGAACCTTCTACTGCTTTTAAACAAGACCATATCTACAAAGACCA 404  
QY 2145 GGATTTGATTAACAATAATAATACACTAATAGAGAATAATGCTTCAGGTCTCTCATCTATAT 2204  
DB 403 GGATTTGATTAACAACAGTATAATACATAATAGAGAATAATGCTTCAGGTCTCTCATCTATAT 344  
QY 2205 TGTGGGTGAGCGTTTATGTACCTGGAGTGGGAATGTGACCAAGAGAGGTCAACAATGAG 2264  
DB 343 TGTGGGAGAACGTTTATGTACCTGGAGTGGGAATGTGACCAAGAGAGGTCAACAATGAG 284  
QY 2265 AGAATATCATCTACTTCTTTCAGTTCGATTCACCCATGCCACACAGTCCCATTCCTCAATTT 2324  
DB 283 AGAGATATCTACTTCTTTCAGTTCGATTCAGCCATGAGCCATGAGTCCCATTCCTCAATTT 224  
QY 2325 ACCTGAGAATGAAATAATGAAACTGGCTTTAGAGAATGTGATTAACAAGTGGCCACATT 2384  
DB 223 ACCTGAGACGAAATAATGAAACTGGCTTTAGAGAATGTGATTAACAAGTGGCCACATT 164  
QY 2385 TAAGAAACAGGTGTATCAGGCCATGGAGTTTATGAACCTAAAGATGAATCACTGGAAGA 2444  
DB 163 TAAGAAACAGGTGTATCAGGCCATGGAGTTTATGAATTAAGATGAATCACTGGAAGA 104  
QY 2445 CTTCAATATGTACTTTTATCATTACTCCAAAACCCAGCAGTAGCAAGGCTGAACATATGCA 2504  
DB 103 CTTCAATATGTACTTTTATCATTACTTAAACACACAGTAGCAAGGCTGAACATATGCA 44  
QY 2505 GAAGAAACGAGAAACAAAGAAAGATGAA 2538  
DB 43 GAAGAAACGAGAAACAAAGAAAGATGAA 10  
RESULT 14  
AK078173  
LOCUS  
DEFINITION AK078173.1 GI:26347022  
Mus musculus adult male olfactory brain cDNA, RIKEN full-length  
enriched library, clone:6430402E23 product:ubiquitin protein ligase  
E3 component n-recognin 1, full insert sequence.  
ACCESSION AK078173  
VERSION AK078173.1  
KEYWORDS HTC; CAP trapper.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1  
Carinci, P. and Hayashizaki, Y.  
TITLE High-efficiency full-length cDNA cloning



JOURNAL  
MEDLINE  
PUBMED

Meth. Enzymol. 303, 19-44 (1999)  
99279253  
10349636

REFERENCE  
AUTHORS

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes  
Genome Res. 10 (10), 1617-1630 (2000)

20499374

JOURNAL  
MEDLINE

11042159

REFERENCE  
AUTHORS

Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuina, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer  
Genome Res. 10 (11), 1757-1771 (2000)

20530913

REFERENCE  
AUTHORS

11076861

JOURNAL  
MEDLINE

Kawai, J., Shingawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nikaido, I., Pesole, G., Quackenbush, J., Schmitt, L. M., Staubli, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bult, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seva, T., Shibata, Y., Storch, K. F., Suzuki, H., Toyooka, K., Wang, K. H., Weitz, C., Whittaker, C., Wilming, L., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohsaki, S. and Hayashizaki, Y.  
Functional annotation of a full-length mouse cDNA collection  
Nature 409 (6821), 685-690 (2001)

21085660

JOURNAL  
MEDLINE

11217851

REFERENCE  
AUTHORS

The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.  
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs  
Nature 420, 563-573 (2002)

6 (bases 1 to 3058)

JOURNAL  
MEDLINE

11042159

REFERENCE  
AUTHORS

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.  
Direct Submission  
Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp).

JOURNAL  
MEDLINE

11042159

REFERENCE  
AUTHORS

COMMENT

URL: <http://genome.gsc.riken.go.jp/>, Tel: 81-45-503-9222, Fax: 81-45-503-9216)  
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN, Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.  
Please visit our web site for further details.  
URL: <http://genome.gsc.riken.go.jp/>  
URL: <http://fantom.gsc.riken.go.jp/>

FEATURES

Location/Qualifiers

1. 3058

source

/organism="Mus musculus"

/mol\_type="mRNA"

/db\_xref="FANTOM\_DB:6430402E23"

/db\_xref="taxon:10090"

/clone="6430402E23"

/sex="male"

/tissue\_type="olfactory brain"

/clone\_lib="RIKEN full-length enriched mouse cDNA library"

/dev\_stage="adult"

<1. 691

/note="unnamed protein product; putative

ubiquitin protein ligase E3 component n-recognin 1

(MGJMGJ:1277977, GI:NM\_009461, evidence: BLASTN, 100%, match=2560)"

/codon\_start=2

/protein\_id="BAC37160.1"

/db\_xref="GI:26347023"

/translation="GVAPPPELFANSAREGESALCSYLSLTNLFLLRQEWDTIRPL

LQWGDPLALKSUKQAVVYPRKNSLIELPESYLLNSHAFRCPSADDERK

HPVLCLFGAILCSQNCQVEVGCACVHALHAGAGVCIFLKTRECRVLLVEG

KARCAYPAPLYDEYGETDPLKRGNPLHLRSRYRKLHLVWQOHCIIIEIARSOETN

QMLFGFNWLL"

3032..3037

/note="putative"

3058

/note="putative"

BASE COUNT 870 a 562 c 632 g 994 t

ORIGIN

Query Match 11.1%; Score 577.4; DB 11; Length 3058;

Best Local Similarity 89.7%; Pred. No. 8.9e-108;

Matches . 620; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

Qy 4515 TGGGTAACTCCGCTGAGGAATGCATACCAATTTCTGCAGAGGAGTACAGTGCCT 4574

1 TGGAGTAGCTCCGCTGAGGAATGTTGCCAATTTCTGCTGAGGAGATTCAGTGCCT 60

Qy 4575 CTGTAGCTATCTATCTTTACCTACAAATTTGTTCTGCTTCCAGGAATATTGGATAC 4634

61 CTGTAGCTATCTATCTTTACCCACAAATTTGTTCTGCTTCCAGGAATATTGGATAC 120

Qy 4635 TGTAAAGCCCTGCTCCAGAGCGGTGTGCAGATCTGCTTACTTAAGTGTTCAGCA 4694

121 CATAAGGCCCTTACTACAGAGGTGTGTGGAGATCTGCTTACCAAGTCTTTGAAGCA 180

Qy 4695 AAAAACAACCCGTGGTCCAGTACCCCTAGAAAAAGAAATAGTTGTATAGAGCTTCCCTGATCA 4754

181 GAAAGTCTGTGGTGCAGTACCCCTAGAAAAAGAAATAGTTGTATAGAGCTTCCCTGATCA 240

Qy 4755 CTATAGCTCCCTCCCTGAATCAAGCTTCTCATTTTCAGGTGCCACGGTCTGCAGATGATGA 4814

241 CTACAGCTGTCTTCTAAATCAGGCTTCTCACATTTAGGTGTCCACGGTCTGCAGATGATGA 300

Qy 4815 GCGAAGCATCTGTCTCTCTGCTTCTGTGGGGCTATATCTTCTCTCAGAACATTTG 4874

301 GCGAAGCATCTGTCTCTCTCTGCTTCTGTGGGGCTATATCTTCTCTCAGAACATTTG 360

Qy 4875 CTGCCAGAAATTTGGAACGGGAGAGTGTGGAGCTTGCATTTTTCACGGCACTTTCACGT 4934

361 TTGGCAAGAAATAGTGAATGGGAGAGGTTGGAGCGTGGCGTTCATTCGCGCTTCATTG 420

```

FEATURES
source
seq primer: p1A-3.
Location/Qualifiers
1. .729
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE: 6404927"
/tissue_type="whole brain"
/dev_stage="embryo 12.5 dpc"
/lab_host="DH10B (T1 phage resistant)"
/clone_lib="NIH_BMAP_FD0"
/note="Organ: Brain; Vector: pYX-Asc; Site_1: EcoR I;
Site_2: Not I; The library was constructed according to
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured mRNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with an
oligo-dT primer containing a Not I site. Double stranded
cDNA was size selected according to mRNA size fraction,
ligated with EcoR I adaptor, digested with Not I, and then
cloned directionally into pYX-Asc vector. The library tag
sequence located between the Not I site and the polyA tail
, is TGAGAGAGCC. This library was created for the
University of Iowa Mouse Brain Molecular Anatomy Project

```

Search completed: September 27, 2003, 13:31:05  
Job time : 6523.54 secs

Search completed: September 27, 2003, 13:31:05  
Job time : 6523.54 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 25, 2003, 14:45:45 ; Search time 104 Seconds  
(without alignments)  
2646.460 Million cell updates/sec

Title: US-09-724-126A-19  
Perfect score: 9141  
Sequence: 1 AMGNWADEAGTEREIS.....EIARSQTNMLFGFNWOLL 1734

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_19Jun03.\*  
1: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*  
24: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9031	98.8	1738	24	ABP58330 Human cell growth,
2	8480	92.8	1757	20	AAW84351 Murine ubiquitin-p
3	8480	92.8	1757	22	AAB31162 Amino acid sequenc
4	4209	46.0	811	22	AAB93464 Human protein sequ
5	3359	36.7	1400	22	AAW78576 Human protein SEQ
6	3356	36.7	1400	22	AAW79560 Human protein SEQ
7	2612.5	28.6	1109	20	AAH02376 Polypeptide identi
8	2210.5	24.2	1829	22	ABB64216 Drosophila melanog
9	2168	23.7	424	23	ABB98104 Human ubiquitin re

10	1716	18.8	333	20	AAW84353 Partial human ubiq
11	1716	18.8	333	22	AAB31163 Amino acid sequenc
12	1071.5	11.7	452	23	ABB90168 Human polypeptide
13	877	9.6	487	22	ABG05917 Novel human diagno
14	790	8.6	258	22	ABG05881 Novel human diagno
15	606.5	6.6	1225	23	ABG93333 Herbicidally activ
16	449	4.9	389	22	ABG06664 Novel human diagno
17	414.5	4.5	203	23	ABB89137 Human polypeptide
18	408.5	4.5	951	24	ABU26663 Human protein modi
19	362	4.0	1456	22	ABB58229 Drosophila melanog
20	360	3.9	153	22	ABH89524 Human immune/haema
21	311	3.4	783	22	ABE58228 Drosophila melanog
22	251	2.7	250	22	AAW25572 Human protein sequ
23	250	2.7	247	21	ABP38334 Human secreted pro
24	240	2.6	81	24	ABP54255 Human lung specifi
25	208	2.3	2230	24	ABU07445 Protein differenti
26	197.5	2.2	2096	21	ABH41592 Human ORFX ORF1356
27	196	2.1	953	22	ABU53071 Intracellular traf
28	189.5	2.1	2048	22	AAW40027 Human polypeptide
29	186	2.0	951	22	ABU53070 Intracellular traf
30	184	2.0	1411	17	AAW02258 Nucleolar/endosoma
31	182	2.0	2442	21	AAW77575 Human cytoskeletal
32	180	2.0	996	22	AAU33755 Staphylococcus aur
33	180	2.0	1009	22	AAU36548 Staphylococcus aur
34	179.5	2.0	2856	24	ABU57575 Mouse lrbA protein
35	178.5	2.0	961	22	ABU53077 Intracellular traf
36	177.5	1.9	2816	22	AAU68572 Human novel cytoki
37	177	1.9	1373	24	ABU11772 Drosophila melanog
38	176	1.9	2013	22	ABE62322 Novel human diagno
39	176	1.9	2017	22	ABG06301 Human polypeptide
40	176	1.9	2888	22	AAW40883 Human 160KB mediat
41	175.5	1.9	1427	12	AAW10534 Novel human diagno
42	175.5	1.9	2633	22	ABG06505 Human polypeptide
43	175.5	1.9	2663	22	AAW39097 Human breast speci
44	175	1.9	2383	23	ABG65631 Restin protein seq
45	174.5	1.9	1392	20	AAW06999

ALIGNMENTS

RESULT 1  
ABP58330  
ID ABP58330 standard; Protein; 1738 AA.  
XX  
AC ABP58330;  
XX  
DT 07-APR-2003 (first entry)  
XX  
DE Human cell growth, differentiation and death protein CGDD-1.  
XX  
KW CGDD-1; cell growth; cell differentiation; cell death; human;  
KW cytostatic; antiarteriosclerotic; hepatotropic; antiinflammatory;  
KW antipsoriatic; antianaemic; ophthalmological; auditory;  
KW anticonvulsant; cerebroprotective; nootropic; neuroprotective;  
KW antiparkinsonian; neuroleptic; tranquilizer; immunosuppressive;  
KW anti-Hiv; antiallergic; antitachymatic; antithyroid; antidiabetic;  
KW dermatological; nephrotropic; antirheumatic; antiarthritic;  
KW antiulcer; vulnerary; virucide; antibacterial; fungicide;  
KW antiparasitic; protozoacide; antihelminthic; antifertility;  
KW gynaecological; ubiquitin protein ligase; enzyme; gene therapy.  
XX  
OS Homo sapiens.  
XX  
PN WO200297032-A2.  
XX  
PD 05-DEC-2002.  
XX  
PF 05-APR-2002; 2002WO-US111152.  
XX  
PR 06-APR-2001; 2001US-282110P.  
PR 11-APR-2001; 2001US-283294P.  
PR 26-APR-2001; 2001US-286820P.

PR 27-APR-2001: 2001US-287228P.  
PR 16-MAY-2001: 2001US-291562P.  
PR 18-MAY-2001: 2001US-291846P.  
PR 25-MAY-2001: 2001US-293727P.  
PR 01-JUN-2001: 2001US-295263P.  
PR 01-JUN-2001: 2001US-295340P.  
PR 15-JAN-2002: 2002US-349705P.  
XX (INCY-) INCYTE GENOMICS INC.  
XX  
XX Azimzai Y, Au-Young JK, Batra S, Baughn MR, Becha SD, Borowsky KJ,  
PI Burford N, Ding L, Elliott VS, Emerling BM, Gandhi AR, Gietzen MJ;  
PI Griffin JA, Hafalia AJA, Honchell CD, Lal PG, Lee SY, Lu DM;  
PI Arvizu CS, Ramkumar J, Reddy R, Sanjanwala MM, Tang YT, Walla NK;  
PI Wang YE, Warren BA, Xu Y, Yang J, Yao MG, Yue H, Zebajadian Y;  
XX  
XX WPI; 2003-140453/13.  
DR N-PSDB; AB224689.  
XX  
PT Novel human proteins associated with cell growth, differentiation and  
PT death, useful for treating, diagnosing or preventing cancer,  
PT developmental, neurological, reproductive or autoimmune/inflammatory  
PT disorders -  
XX  
XX Claim 1; Page 183-187; 238pp; English.  
PS  
XX The present sequence is the protein sequence of human CGDD-1, a  
CC novel protein associated with cell growth, differentiation and  
CC death. The sequence is predicted from Incyte clone 1351608CB1,  
CC which was isolated from a paraneoplastic tumour tissue CDNA  
CC library. Structural features establish the protein as being  
CC associated with cell growth, differentiation and death, with  
CC further evidence suggesting it to be a ubiquitin protein ligase.  
CC The invention is based on novel human CGDD-1 to -21 proteins (see  
CC ABP58330-50), the polynucleotides encoding them (see AB224689-709),  
CC and to the use of these for the diagnosis, treatment or prevention  
CC of cell proliferative disorders including cancer, developmental  
CC disorders, neurological disorders, autoimmune disorders,  
CC reproductive disorders, and disorders of the placenta, and in the  
CC assessment of the effects of exogenous compounds on the activity  
CC and expression of proteins and nucleic acids associated with cell  
CC growth, differentiation and death.  
XX  
SQ Sequence 1738 AA;  
Query Match 98.8%; Score 9031; DB 24; Length 1738;  
Best Local Similarity 98.8%; Pred. No. 0;  
Matches 1717; Conservative 0; Mismatches 1; Indels 20; Gaps 1;  
QY 17 MEISALPOTPORLASWMDQVDYTAFLHLLAQLVPEIYFAEMDPLEKQESVQMSIF 76  
DB 1 MEISALPOTPORLASWMDQVDYTAFLHLLAQLVPEIYFAEMDPLEKQESVQMSIF 60  
QY 77 TPLEWLFGEDDPICLEKLKHSAGFOLCGRVFKSGGTTYSRCRCAIDPTCVLCMDCFQDS 136  
DB 61 TPLEWLFGEDDPICLEKLKHSAGFOLCGRVFKSGGTTYSRCRCAIDPTCVLCMDCFQDS 120  
QY 137 VHKNHRYKMHSTGGGFCDCGDTAEWKTPFCVNHPEPGRAGTIKENSRCPLNEEIVQAR 196  
DB 121 VHKNHRYKMHSTGGGFCDCGDTAEWKTPFCVNHPEPGRAGTIKENSRCPLNEEIVQAR 180  
QY 197 KIPPSVIKYVEMTWEEKEKLPPELQIREKNERYCYVLFNDEHSHYDHYISLQALDC 256  
DB 181 KIPPSVIKYVEMTWEEKEKLPPELQIREKNERYCYVLFNDEHSHYDHYISLQALDC 240  
QY 257 ELAEALHTTAIDKGRRAVKAGAYAAQCEAKEDIKSHSENSVQHPHLHVEVLHSEIMAHQ 316  
DB 241 ELAEALHTTAIDKGRRAVKAGAYAAQCEAKEDIKSHSENSVQHPHLHVEVLHSEIMAHQ 300  
QY 317 KFAALRGSMWNKIMSYSSDFRQIFCOACLRPEPDSENPCILSRMLWDKAKLYKGARKILH 376  
DB 301 KFAALRGSMWNKIMSYSSDFRQIFCOACLRPEPDSENPCILSRMLWDKAKLYKGARKILH 360

QY 377 ELIFSSFFMEMEYKKLFAMEFVKYKOLQKEYISDDHDSRISITALSVMQFTVPTLARHL 436  
DB 361 ELIFSSFFMEMEYKKLFAMEFVKYKOLQKEYISDDHDSRISITALSVMQFTVPTLARHL 420  
QY 437 IEEQNVISVITETLLEVLPEYLDNRNKNFQGYSDQKLGRIYVAYICDLKYILISKPTIWT 496  
DB 421 IEEQNVISVITETLLEVLPEYLDNRNKNFQGYSDQKLGRIYVAYICDLKYILISKPTIWT 480  
QY 497 ERRLMOPLEGFRSFLKILTQMGMEETRRQVGHIEVDPDWEAAIAIOMQLKNILLMPQE 556  
DB 481 ERRLMOPLEGFRSFLKILTQMGMEETRRQVGHIEVDPDWEAAIAIOMQLKNILLMPQE 540  
QY 557 WCACDELLAVAYKECHKAVMRCSTSFISSTKTVQSCGSHLETKSVRSVSDLSIHLPL 616  
DB 541 WCACDELLAVAYKECHKAVMRCSTSFISSTKTVQSCGSHLETKSVRSVSDLSIHLPL 600  
QY 617 SRTLGLHVRSLRSLGAVSRLEHFEVFEVLEVEYPLRCLVLVAQVVAEMWRNGLSL 676  
DB 601 SRTLGLHVRSLRSLGAVSRLEHFEVFEVLEVEYPLRCLVLVAQVVAEMWRNGLSL 660  
QY 677 ISOVFFYQDYKCREMYDKDIIMLQIGASLMDPNKFLLLVLRQYELAEAFNKTISTKDDQ 736  
DB 661 ISOVFFYQDYKCREMYDKDIIMLQIGASLMDPNKFLLLVLRQYELAEAFNKTISTKDDQ 720  
QY 737 LIKOYNTLIEEMLOVLIIYIGERYVPGVGNVTKEEVTMRETIHLLCTEPMHPSATAKNLP 796  
DB 721 LIKOYNTLIEEMLOVLIIYIGERYVPGVGNVTKEEVTMRETIHLLCTEPMHPSATAKNLP 780  
QY 797 ENENNETGLENVINKVATFKKPGVSGHGYELKDESLEKDFNMVYHYHYSKTOHSAEHMOK 856  
DB 781 ENENNETGLENVINKVATFKKPGVSGHGYELKDESLEKDFNMVYHYHYSKTOHSAEHMOK 840  
QY 857 KRRKQENKDEALPPPPPEFCPAFSKVINLNCDDIMYILRTVFERAIDTDSNLTWEGML 916  
DB 841 KRRKQENKDEALPPPPPEFCPAFSKVINLNCDDIMYILRTVFERAIDTDSNLTWEGML 900  
QY 917 QMAFHIALGLLEKQOLQKAPAEVEVDFYHFKASRLGSSAMNTOMLLEKLGIPQLEGQ 976  
DB 901 QMAFHIALGLLEKQOLQKAPAEVEVDFYHFKASRLGSSAMNTOMLLEKLGIPQLEGQ 960  
QY 977 KDMITWLOMFDTVKRLREKSCLLVATTSGSESINKDEITHDKAEKRAEAAARLRQ 1036  
DB 961 KDMITWLOMFDTVKRLREKSCLLVATTSGSESINKDEITHDKAEKRAEAAARLRQ 1020  
QY 1037 KIMAQMSALQKNFTETHKLMYDNTSEMPGKEDSIMEESTPAVSYSRIALGPKRGPSVT 1096  
DB 1021 KIMAQMSALQKNFTETHKLMYDNTSEMPGKEDSIMEESTPAVSYSRIALGPKRGPSVT 1080  
QY 1097 EKEVLTICLQCEOEVEKIENNAMVLSACVQKSTALTQHRGKPIELSGEALDPLFMDPDLA 1156  
DB 1081 EKEVLTICLQCEOEVEKIENNAMVLSACVQKSTALTQHRGKPIELSGEALDPLFMDPDLA 1140  
QY 1157 YGVTGSCGHVMHACVQKFEAVOLSSQQRHVDLFDLESGEYLCPLCKSLCNTVPIPI 1216  
DB 1141 YGVTGSCGHVMHACVQKFEAVOLSSQQRHVDLFDLESGEYLCPLCKSLCNTVPIPI 1200  
QY 1217 PLOPKNTSENADALAQLLTARWIQTVLARISGYNIRHAKGENPIPIFFNOGMDSTLE 1276  
DB 1201 PLOPKNTSENADALAQLLTARWIQTVLARISGYNIRHAKGENPIPIFFNOGMDSTLE 1260  
QY 1277 FHSILSGVSESSIKYSNIKEMVILFATTIYRIGLKVPPDERDPRVPMLTWSTCAFTTQA 1336  
DB 1261 FHSILSGVSESSIKYSNIKEMVILFATTIYRIGLKVPPDERDPRVPMLTWSTCAFTTQA 1320  
QY 1337 IENLLGDEGKPLFGALONROHNGKALMQFAVORITCPQVLIQKHLVRLSVLPNPKS 1396  
DB 1321 IENLLGDEGKPLFGALONROHNGKALMQFAVORITCPQVLIQKHLVRLSVLPNPKS 1380  
QY 1397 EDTPCLLSIDLPHVLVGAFLAPPSLYWDDPDVLOPSSVSSSYNHLFLPHLTMAHMLQTL 1456  
DB 1381 EDTPCLLSIDLPHVLVGAFLAPPSLYWDDPDVLOPSSVSSSYNHLFLPHLTMAHMLQTL 1440  
QY 1457 LTVDTGLPLAQVQEDSEEAHSASSFFAEISQYTSIGSGICDDIPGWLYWLSLKNIGTIPYLR 1516

```
Db 1441 LTVDGLPLAQVQDESEAHSSAFELISQYTSIGCDIPGWLWLSKNGITPYLRAC 1500
QY 1517 AALFFHYLLGVTTPPELHTNSAEGEYSALCSYLSLPTNLFLLFQYWDTVRPLLRRCAD 1576
Db 1501 AALFFHYLLGVTTPPELHTNSAEGEYSALCSYLSLPTNLFLLFQYWDTVRPLLRWCAD 1560
QY 1577 PALLNCLKQKNTVVRYPRKNSLIELPDDYSCLLNQASHFRCPRSADDERKHPVLCLFCG 1636
Db 1561 PALLNCLKQKNTVVRYPRKNSLIELPDDYSCLLNQASHFRCPRSADDERKHPVLCLFCG 1620
QY 1637 AILCSNCCOEIVNGEEVGACIFHALHC-----KARGCAYPAPY 1676
Db 1621 AILCSNCCOEIVNGEEVGACIFHALHCAGVCIFLIRECRVVLVEGKARGCAYPAPY 1680
QY 1677 LDEYGETDPLGRGNPLHLRSERYRKLHLVWQOHCIIIEIARSQETNOMLFQFNWOLL 1734
Db 1681 LDEYGETDPLGRGNPLHLRSERYRKLHLVWQOHCIIIEIARSQETNOMLFQFNWOLL 1738

RESULT 2
AAW84351
ID AAW84351 standard; Protein; 1757 AA.
AC AAW84351;
XX
DT 25-MAR-1999 (first entry)
XX
DE Murine ubiquitin-protein ligase Ubr1.
XX
KW Ubiquitin-protein ligase; Ubr1; mouse; ubiquitinylation; degradation;
N-end rule pathway; stress-related muscle wasting; inhibitor; screen.
XX
OS Mus sp.
XX
PN US5861312-A.
XX
PD 19-JAN-1999.
XX
PF 02-DEC-1997; 97US-0982956.
XX
PR 02-DEC-1997; 97US-0982956.
XX
PA (CALY ) CALIFORNIA INST OF TECHNOLOGY.
XX
PI Kwon YT, Varshavsky A;
XX
DR WPI; 1999-130395/11.
DR N-PSDB; AAV99308.
XX
PT Mouse and human Ubr1 cDNA - useful for producing recombinant Ubr1
polypeptides
XX
PS Disclosure; Columns 15-28; 18pp; English.
XX
CC The present sequence represents a ubiquitin-protein ligase called Ubr1.
CC The Ubr1 enzymes are involved in protein ubiquitinylation and
CC ultimate degradation through the N-end rule pathway and have been
CC linked to stress-related muscle wasting. Recombinant Ubr1 polypeptides
CC can be used to screen for inhibitors of muscle wasting when this is
XX associated with the N-end rule pathway.
SQ Sequence 1757 AA;

Query Match 92.8%; Score 8480; DB 20; Length 1757;
Best Local Similarity 91.0%; Pred. No. 0;
Matches 1598; Conservative 64; Mismatches 67; Indels 28; Gaps 4;

QY 6 MADEAGCTERMEISAEPLQTPORLASWWDQVDFTAFHLHQAOLVPEIFAEMDPDLE 65
Db 1 MADEMDGAEKMDVSPPEPLAPQRPASWWDQVDFTAFHLHQAOLVPEIFAEMDPDLE 60
QY 66 KOESVQMSIITPLEWYLFGEPPDICLEKLSHGAFLCGRVKSGSETTYSRCDAIDPT 125
```

```
Db 61 KOESVQMSIITPLEWYLFGEPPDICLEKLSHGAFLCGRVKSGSETTYSRCDAIDPT 120
QY 126 CVLCMDFQDSVHKHNRKMKHTSTGGGFCDCGDTFAMKTPFCVNHPEGRAGTIKENSRC 185
Db 121 CVLCMDFQDSVHKHNRKMKHTSTGGGFCDCGDTFAMKTPFCVNHPEGRAGTIKESLHC 180
QY 186 PLNEEVIVQARKTIPSPVIKYVEMTWEEKEKLPPELQIREKNERYCVLENDEHSHSDH 245
Db 181 PLNEEVIAQARRIFPSPVIKYVEMTWEEKEKLPPELQIREKNERYCVLENDEHSHSDH 240
QY 246 VIYSLOALDCELAEAQLHTTADKEGRRVAVKAGAYAACQAEKEDIKSHSENVSOHPLHV 305
Db 241 VIYSLOALDCELAEAQLHTTADKEGRRVAVKAGAYATCQAEKEDIKSHSENVSOHPLHV 300
QY 306 EVLHSEITMAHOKFALRLGSMWNKIMSYSSDFQIFCOACLRDEEDSNPCILISRLMLWDA 365
Db 301 EVLHSEVMAHOKFALRLGSMWNKIMSYSSDFQIFCOACLRDEEDSNPCILISRLMLWDA 360
QY 366 KLYKGARKILHELIFSSFFMEMEYKLLFAMEFYKYKQLOKEYISDDHDSISITALSVQ 425
Db 361 KLYKGARKILHELIFSSFFMEMEYKLLFAMEFYKYKQLOKEYISDDHDSISITALSVQ 420
QY 426 MFTVPTLARHLIBEQNVISVITETLLEVLPEYLDNRNKNFQGYSDQKLGRYAVICDLK 485
Db 421 MFTVPTLARHLIBEQNVISVITETLLEVLPEYLDNRNKNFQGYSDQKLGRYAVICDLK 480
QY 486 YLISKPTIWTETLRMOFLGFRSFLKILTCMOGMEIRQVGOHIEVDPDWEAAIATQM 545
Db 481 YLISKPTIWTETLRMOFLGFRSFLKILTCMOGMEIRQVGOHIEVDPDWEAAIATQM 540
QY 546 QLKNIILLMFQEWACDEELLLVAYKECHKAVMRCTSFISSSKTVVSCGSHSLTKSYRV 605
Db 541 QLKNIILLMFQEWACDEELLLVAYKECHKAVMRCTSFISSSKTVVSCGSHSLTKSYRV 600
QY 606 SEDLVSIHPLSRTLAGLVRLSRLGAVSRHLEFVSFEDFQVEVLVEYPLRCLVLVAQVV 665
Db 601 SEDLVSIHPLSRTLAGLVRLSRLGAVSRHLEFVSFEDFQVEVLVEYPLRCLVLVAQVV 660
QY 666 AEMWRNGLSLISQVFIYQDVCKREEMDKDILMLOIGASIMDPNKFLLVLQRYELAE 725
Db 661 AEMWRNGLSLISQVFIYQDVCKREEMDKDILMLOIGASIMDPNKFLLVLQRYELAE 720
QY 726 FNKTIKQDQDLIKQYNTLIEEMLOVLIYIGERYVPGVGNVKEEYTMREIHLICIEP 785
Db 721 FNKTIKQDQDLIKQYNTLIEEMLOVLIYIGERYVPGVGNVKEEYTMREIHLICIEP 780
QY 786 MPHSAITAKNLPENENNETGLENVINKVATFKKPGVSGHGVYELKDESLKDFNMFFYHYSK 845
Db 781 MPHSAITAKNLPENENNETGLENVINKVATFKKPGVSGHGVYELKDESLKDFNMFFYHYSK 840
QY 846 TQHSKAHEMOKKRRKQENKDEALPPPPPECFPAFSKVINLLNCIDIMWYILRTVFERAID 905
Db 841 TQHSKAHEMOKKRRKQENKDEALPPPPPECFPAFSKVINLLNCIDIMWYILRTVFERAID 900
QY 906 TQSNLWTEGMLQAFHILALGLEEKQOLKAPDEEVTDFYHKASRLGSSAM--NIQM 962
Db 901 TQSNLWTEGMLQAFHILALGLEEKQOLKAPDEEVTDFYHKASRLGSSAMNAQIOM 960
QY 963 LLEKLGIPQLEGQKDMITWILQMFDTVKRLREKSLIVATTSGSEIKNDEIHTHDEKA 1022
Db 961 LLEKLGIPQLEGQKDMITWILQMFDTVKRLREKSLIVATTSGSEIKNDEIHTHDEKA 1020
QY 1023 ERKRKAARLHRQKIQAQMSALQKNFIETHKLMYDNTSEMPGKEDSIMEESTPVSVDY 1082
Db 1021 ERKRKAARLHRQKIQAQMSALQKNFIETHKLMYDNTSEMPGKEDSIMEESTPVSVDY 1080
QY 1083 SRTALGPKRGPSTYKTEVELTCILCOEVEQVKIENNAVLSACVQKSTALTOHRKPKTELS 1142
Db 1081 SRTALGPKRGPSTYKTEVELTCILCOEVEQVKIENNAVLSACVQKSTALTOHRKPKTELS 1140
QY 1143 GEALDPLFMDPDLAYGTYTGSCGRVMAHVCQKYFEAVQLSSQQRHVDLFDLESGBYLC 1202
```

Db 1141 GETLDPLFMDPLAHGTYTSCGHVMHVCWQKYFEAVQLSSQRIHVDLFDLEGEYLC 1200  
 Qy 1203 PLCKSLCNVTIPIIPIQOKINSENADALQALLTLARWQTIVLARIYSGYNIRAKGNP- 1261  
 Db 1201 PLCKSLCNVTIPIIPIQOKINSENADALQALLTLARWQTIVLARIYSGYNIRAKGNP 1260  
 Qy 1262 IPIFFNOGMDSTLEFHSILSGVSSIKYSIKEMVILFATYIRIGLKVDPDERDP 1321  
 Db 1261 VPVLFNOGMDSTLEFHSILSGVSSIKYSIKEMVILFATYIRIGLKVDPDERDP 1320  
 Qy 1322 VPMLTWSTCAFTIQAIENTLNGDGKPLFGALQNRQINGLKALMQFAVAQAITCPQVLIQ 1381  
 Db 1321 VPMWSTCAFTIQAIENTLNGDGKPLFGALQNRQINGLKALMQFAVAQAITCPQVLIH 1380  
 Qy 1382 HLVRLLSVLPNTKSEDTCLLSIDILFHLVGLAVLAPPSLYWDDPDVQLPSSVSSYNHL 1441  
 Db 1381 HLRLLSVLPNTKSEDTCLLSIDILFHLVGLAVLAPPSLYWDDPDVQLPSSVSSYNHL 1440  
 Qy 1442 YLFHLITMAHMLQILLTVDI---GLPLAQVQEDSEAHSAFFFAEISQYTSIGICDI 1497  
 Db 1441 YLFHLITMAHMLQILLTDTDLSGPPPLAEGEEDSEARCAFAFFVEVSQHTDGLTCCGA 1500  
 Qy 1498 PGYLVWSLKNGITPYLRCAALFFHYLLGVTPPEELHTNSAEGEYSAICLSYLSLPTNLF 1557  
 Db 1501 PGYLVWSLKNGITPYLRCAALFFHYLLGVTPPEELHTNSAEGEYSAICLSYLSLPTNLF 1560  
 Qy 1558 LFOEYWDTRVPLQRRCADPALLCNLCIKQNTVYPRKRNSLTIELPDDYSCLLNQASHFR 1617  
 Db 1561 LFOEYWDTRVPLQRRCADPALLCNLCIKQNTVYPRKRNSLTIELPDDYSCLLNQASHFR 1620  
 Qy 1618 CPRSADDERKHPVLCFLCGAILCSQNICQOEIYNGEVEGACIFALHC----- 1665  
 Db 1621 CPRSADDERKHPVLCFLCGAILCSQNICQOEIYNGEVEGACIFALHC----- 1680  
 Qy 1666 -----KARGCAYPAPYLDEYGETDPLKRGKGNPLHLRSRYRKHLVWQHCIIIEIA 1717  
 Db 1681 CRVVLVEGKARGCAYPAPYLDEYGETDPLKRGKGNPLHLRSRYRKHLVWQHCIIIEIA 1740  
 Qy 1718 RSQETNOMLFGFNWQLL 1734  
 Db 1741 RSQETNOMLFGFNWQLL 1757

## RESULT 3

AAB31162  
 ID AAB31162 standard; Protein; 1757 AA.

AC AAB31162;

XX 02-APR-2001 (first entry)

DT Amino acid sequence of Mouse Ubr1 protein.

DE Ubr1; E3-type protein; ubiquitin system; ubiquitin-protein ligase;  
 KW N-end rule pathway; intracellular pathogen; Lysteria monocytogenes;  
 KW Yersinia enterocolitica; muscle wasting; infection.

XX Mus sp.

XX US6159732-A.

XX 12-DEC-2000.

XX 11-JAN-1999; 99US-0228317.

XX 02-DEC-1997; 97US-0982956.

PR (CALY ) CALIFORNIA INST OF TECHNOLOGY.

XX Kwon YT, Varshavsky A;

XX WPI; 2001-090278/10.

DR N-PSDB; AAC86933.

XX Inhibiting the N-end rule pathway in mammalian cells for treating  
 PT infections and various diseases associated with muscle tissue wasting,  
 PT by inhibiting the expression of Ubr1 gene -  
 XX  
 PS Example; Column 15-28; 18pp; English.

XX The present sequence represents a murine Ubr1 enzyme. Ubr1 is an E3-type  
 CC protein of the ubiquitin system. Specifically, it is a ubiquitin-protein  
 CC ligase. The enzyme is specific for destabilising residues exposed at  
 CC the N-terminus of protein substrates. Inhibition of the expression of  
 CC Ubr1 gene in a cell results in inhibition of the N-end rule pathway.  
 CC The method is used for treatment of mammalian cells infected with an  
 CC intracellular pathogen, e.g. Lysteria monocytogenes or Yersinia  
 CC enterocolitica. Inhibition of N-end rule pathway is also useful for  
 CC treating various diseases associated with wasting of muscle tissue and  
 CC infections.

XX SQ Sequence 1757 AA;

Query Match 92.8%; Score 8480; DB 22; Length 1757;  
 Best Local Similarity 91.0%; Pred. No. 0;  
 Matches 1598; Conservative 64; Mismatches 67; Indels 28; Gaps 4;

Qy 6 MADEAGCTERMEISAEELPQTPORLASWMDQVDYTFATLHHLAQLVPEIYFAEMDPDLE 65  
 Db 1 MADEEMDGAERMDVSPEPLAPQRPASWMDQVDYTFATLHHLAQLVPEIYFAEMDPDLE 60  
 Qy 66 KOESVQMSIITPLEWYLFEGDDPDICLEKLSHGAFLCGRVRKSGETTYSCRDCAIDPT 125  
 Db 61 KOESVQMSIITPLEWYLFEGDDPDICLEKLSHGAFLCGRVRKSGETTYSCRDCAIDPT 120  
 Qy 126 CVLQCMDFQDSVHKNHRYKMHSTGGGFCDCGDTGEMTKTPFCVNHPEFRAGTIKENSRC 185  
 Db 121 CVLQCMDFQDSVHKNHRYKMHSTGGGFCDCGDTGEMTKTPFCVNHPEFRAGTIKENSRC 180  
 Qy 186 PLNEEVIQARKIPSPVYVYVEMTIWEEELPELOIREKNERYCYVLFNDEHSHVDH 245  
 Db 181 PLNEEVIQARKIPSPVYVYVEMTIWEEELPELOIREKNERYCYVLFNDEHSHVDH 240  
 Qy 246 VIYSLQALDCELAELAAQLHTTAIDKEGRRAVKAGAYACQAEKEDIKSHSENVSOHPH 305  
 Db 241 VIYSLQALDCELAELAAQLHTTAIDKEGRRAVKAGAYATCQAEKEDIKSHSENVSOHPH 300  
 Qy 306 EVLHSEIMAHQKFAIRLGSWMNKIMSYSDFRQIFCQACLRPEEPSENPCILSRMLMDA 365  
 Db 301 EVLHSEIMAHQKFAIRLGSWMNKIMSYSDFRQIFCQACLRPEEPSENPCILSRMLMDA 360  
 Qy 366 KLYKGARKILHELIFSSFFMEMEYKKLFAMEFVKYKQLOKEYISDDHRSISITALSVO 425  
 Db 361 KLYKGARKILHELIFSSFFMEMEYKKLFAMEFVKYKQLOKEYISDDHRSISITALSVO 420  
 Qy 426 MFTVPTLARHLIEQNVISVITETLLEVLPEYLDNRNKNFNFGYSQDKLGRVYAVICDLK 485  
 Db 421 MFTVPTLARHLIEQNVISVITETLLEVLPEYLDNRNKNFNFGYSQDKLGRVYAVICDLK 480  
 Qy 486 YILISKPTITWELRMQFLEGFRSKLTICMQGMEETRRQYQGHIEVDPDWEAAIAIQM 545  
 Db 481 YILISKPTITWELRMQFLEGFRSKLTICMQGMEETRRQYQGHIEVDPDWEAAIAIQM 540  
 Qy 546 QLNKILLMFQEWACDEBELLVAYKECHKAVNRCSTSFISSTKTVVQSGCHSLETKSYRV 605  
 Db 541 QLNKILLMFQEWACDEBELLVAYKECHKAVNRCSTSFISSTKTVVQSGCHSLETKSYRV 600  
 Qy 606 SEDLVS IHLPLSRTLAGLHVRSLRGAVSRLEHFVSEDFQVEVLVEYPLRCLVLVAQVY 665  
 Db 601 SEDLVS IHLPLSRTLAGLHVRSLRGAVSRLEHFVSEDFQVEVLVEYPLRCLVLVAQVY 660  
 Qy 666 AEMWRRNGLSLISQVFFYQDVKCREMYDKDITIMLQIGASLMDPNKFLLLVLRQYELABA 725  
 Db 661 AEMWRRNGLSLISQVFFYQDVKCREMYDKDITIMLQIGASLMDPNKFLLLVLRQYELABA 720  
 Qy 726 FNKTIKTDQDLIKOYNLTLEMLQVLIIYIGERVYVPGVGNVTKSEVTMREIHLICIEP 785

Db 721 FNKTISTKQDQILKQYNLTLEEMQLVLIYIGERYVPGVNTREEVIMREITHLLCIEP 780  
QY 786 MPHSIAIAKNLENENNETGLENINKVATFKPGVSGHGYVELKDESKDFNMYYHYSK 845  
Db 781 MPHSIAIARNLENENNETGLENINKVATFKPGVSGHGYVELKDESKDFNMYYHYSK 840  
QY 846 TOHSAKAEHMKRRKQENKDEALPPPPPEPCPAFSKVINLLNCDDIMMYILRTVFERAID 905  
Db 841 TOHSAKAEHMKRRKQENKDEALPPPPPEPCPAFSKVINLLNCDDIMMYILRTVFERAID 900  
QY 906 TDSNLWTEGLQMAFHIALGLLEKEKQLOKAPAEVEVTFDFYHKASRLGSSAM---NIQM 962  
Db 901 TESNLWTEGLQMAFHIALGLLEKEKQLOKAPAEVEVTFDFYHKASRLGSSAMNAQNTQM 960  
QY 963 LLEKLKIPQLEGQKDMITWILQMFDTVKRUREKSLIVATTSGESIKNDEITHDKEKA 1022  
Db 961 LLEKLKIPQLEGQKDMITWILQMFDTVKRUREKSLIVATTSGESIKNDEITHDKEKA 1020  
QY 1023 ERKEKAEARLHRQKIMAQNSALQKNFTETHKLMYDNTSEMPGKREDSIMEESTPAVSDY 1082  
Db 1021 ERKEKAEARLHRQKIMAQNSALQKNFTETHKLMYDNTSEMPGKREDSIMEESTPAVSEA 1080  
QY 1083 SRIALGPKRPSVTEKEVLTICLQEOEVEKIENNAMVLSACVOKSTALTOHRGKPIELS 1142  
Db 1081 SRIALGPKRPSVTEKEVLTICLQEOEVEKIENNAMVLSACVOKSTALTOHRGKPIELS 1140  
QY 1143 GEALDPLFMDPLDLAGTYTGSCGHVMHVCWQKYFEAVQLSSQQRHVDLFDLESSEYLC 1202  
Db 1141 GETLDPLFMDPLDLAGTYTGSCGHVMHVCWQKYFEAVQLSSQQRHVDLFDLESSEYLC 1200  
QY 1203 PLCKSLCNTVPIPILOPKINSNADALQALLLARIQVILARISGYNIRHAKGNP- 1261  
Db 1201 PLCKSLCNTVPIPILOPKINSNADALQALLLARIQVILARISGYNIRHAKGEAPA 1260  
QY 1262 IPIFFNMGMDSTLEFHSILSFGVESSTIKYSNISKEMVILPATIYIRIGLVKVPDPDR 1321  
Db 1261 IPIFFNMGMDSTLEFHSILSFGVESSTIKYSNISKEMVILPATIYIRIGLVKVPDPDR 1320  
QY 1322 VPMLTWSTCAFTIQAIEINLLGDEGKPLFGALQNRQHGKALMOFAVAQRITCPOVLIOK 1381  
Db 1321 VPMLTWSTCAFTIQAIEINLLGDEGKPLFGALQNRQHGKALMOFAVAQRATCPOVLIOK 1380  
QY 1382 HLVRLLSVLPNIKSEDTPCLLSIDLHVLVGAFLAPSLVWDVDPDLQPSVSSSYNHL 1441  
Db 1381 HLVRLLSVLPNIKSEDTPCLLSIDLHVLVGAFLAPSLVWDVDPDLQPSVSSSYNHL 1440  
QY 1442 YLFLHITMAHMLQILLTVDOT---GLPLAQVQEDSEEAHSAFSAEISQVTSIGCDI 1497  
Db 1441 YLFLHITMAHMLQILLTVDOT---GLPLAQVQEDSEEAHSAFSAEISQVTSIGCDI 1490  
QY 1498 PGWYLMVSLKNGITPYLRCAALFFHYLLGVTPPELHTNSAEGVYSA LCSYLSLPTNLFL 1557  
Db 1501 PGWYLMVSLKNGITPYLRCAALFFHYLLGVTPPELHTNSAEGVYSA LCSYLSLPTNLFL 1560  
QY 1558 LFOEYWDVTRPLQRCADPALLNCLKQNTKNTVRYPRKRNSLIELPDDYSCLLNQASHFR 1617  
Db 1561 LFOEYWDVTRPLQRCADPALLNCLKQNTKNTVRYPRKRNSLIELPDDYSCLLNQASHFR 1620  
QY 1618 CPRSADDERKHPVLCFGAILCSQNTCCQBIIVGEEVGACIFIALHC----- 1665  
Db 1621 CPRSADDERKHPVLCFGAILCSQNTCCQBIIVGEEVGACIFIALHC----- 1660  
QY 1666 -----KARGCAYPAPYLDXEYGETDPLGRGNPLHLRSERYRKLHLVWQOHCIIIEITA 1717  
Db 1681 CRVVLVEGKARGCAYPAPYLDXEYGETDPLGRGNPLHLRSERYRKLHLVWQOHCIIIEITA 1740  
QY 1718 RSQETNQMLFGFNWQLL 1734  
Db 1741 RSQETNQMLFGFNWQLL 1757

AAB93464  
ID AAB93464 standard; Protein; 811 AA.  
XX  
AC AAB93464;  
XX  
DT 26-JUN-2001 (first entry)  
XX  
XX Human protein sequence SEQ ID NO:12732.  
DE Human; primer; detection; diagnosis; antisense therapy; gene therapy.  
KW Homo sapiens.  
XX  
OS EP1074617-A2.  
PN  
XX  
PD 07-FEB-2001.  
XX  
PF 28-JUL-2000; 2000EP-0116126.  
XX  
PR 29-JUL-1999; 99JP-0248036.  
PR 27-AUG-1999; 99JP-0300253.  
PR 11-JAN-2000; 2000JP-0118776.  
PR 02-MAY-2000; 2000JP-0183767.  
PR 09-JUN-2000; 2000JP-0241899.  
XX  
XX (HELI-) HELIX RES INST.  
XX  
PI Ota T, Isoqai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX  
XX WPI; 2001-318749/34.  
XX  
PT Primer sets for synthesizing polynucleotides, particularly the 5602  
PT full-length cDNAs defined in the specification, and for the detection  
PT and/or diagnosis of the abnormality of the proteins encoded by the  
PT full-length cDNAs -  
XX  
PS Claim 8; SEQ ID 12732; 2537pp + CD ROM; English.  
XX  
CC The present invention describes primer sets for synthesizing 5602  
CC full-length cDNAs defined in the specification. Where a primer set  
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
CC to the complementary strand of a polynucleotide which comprises one of  
CC the 5602 nucleotide sequences defined in the specification, where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in  
CC the specification. The primer sets can be used in antisense therapy and  
CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
CC represent oligonucleotides, all of which are used in the exemplification  
CC of the present invention.  
XX  
SQ Sequence 811 AA;

Query Match 46.0%; Score 4209; DB 22; Length 811;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 811; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 209 MTIWEERKELPPELOIREKNRYCYVLFNDEHSHYDHVYISLORALDCELAEOHHTAI 268  
Db 1 MTIWEERKELPPELOIREKNRYCYVLFNDEHSHYDHVYISLORALDCELAEOHHTAI 60  
QY 269 DKGRRAVAGAYAAQCAKEADIKSHSENVSQHPLHVEVLHSETMAHQKFAALGLGSMNK 328



```
Db 61 DREGRAVAGAYAAQAEKEDIKSHSENVSHPLHVEVLHSEIMAHQFALRGSGMNNK 120
QY 329 IMSYSSDFRQIFCQACLRPEPSENCLISRLMLDAKLYKGARKILHELIFSSFFMEME 388
Db 121 IMSYSSDFRQIFCQACLRPEPSENCLISRLMLDAKLYKGARKILHELIFSSFFMEME 180
QY 389 YKKLFAMEFVKYKQKQKEYISDDHRSISITALSVMQETVPTPLARHLIEEQNVISITE 448
Db 181 YKKLFAMEFVKYKQKQKEYISDDHRSISITALSVMQETVPTPLARHLIEEQNVISITE 240
QY 449 TLLEVLPEYLDNRNKNFQGYSDKLGRYAVYICDLKYILISKPTTWTERLRNQFLEGFR 508
Db 241 TLLEVLPEYLDNRNKNFQGYSDKLGRYAVYICDLKYILISKPTTWTERLRNQFLEGFR 300
QY 509 SFKLILTCMQGMEIEIRROVQGHIEVDPDWEAAIAIQMLKNILLMFQEWACACDEELLVA 568
Db 301 SFKLILTCMQGMEIEIRROVQGHIEVDPDWEAAIAIQMLKNILLMFQEWACACDEELLVA 360
QY 569 YKECHKAVMRCSTSFSSSKTVVQSCGHSLETYSYRVSEDLVSIHLPLSRTLAGLHVRLS 628
Db 361 YKECHKAVMRCSTSFSSSKTVVQSCGHSLETYSYRVSEDLVSIHLPLSRTLAGLHVRLS 420
QY 629 RLGAVSRLEHFEVSFDFQVEVLVEYPLRCLVLVAQVVAEMWRRNGLSLISQVFFYQDVKC 688
Db 421 RLGAVSRLEHFEVSFDFQVEVLVEYPLRCLVLVAQVVAEMWRRNGLSLISQVFFYQDVKC 480
QY 689 REEMYDKDIIMLIQIGASLMDPNKFLLLVLQRYELAEAFNKTISTKQDQLIKQYNTLIEEM 748
Db 481 REEMYDKDIIMLIQIGASLMDPNKFLLLVLQRYELAEAFNKTISTKQDQLIKQYNTLIEEM 540
QY 749 LQVLIYIVGERYVPGVGNVTKKEVTMRIEIHLLCIEPMPHSAIAKNLPENNETGLENY 808
Db 541 LQVLIYIVGERYVPGVGNVTKKEVTMRIEIHLLCIEPMPHSAIAKNLPENNETGLENY 600
QY 809 INKVATFKPGVSGHGVYELKDESLKDFNMVYFHYSKTQHSKAEHMQKRRKQENKDEAL 868
Db 601 INKVATFKPGVSGHGVYELKDESLKDFNMVYFHYSKTQHSKAEHMQKRRKQENKDEAL 660
QY 869 PPPPPFPCPAFSKVINLLNCDIMVILRTVFERRADTDSNLTGEMQWAFHILALGLL 928
Db 661 PPPPPFPCPAFSKVINLLNCDIMVILRTVFERRADTDSNLTGEMQWAFHILALGLL 720
QY 929 EEKQQLQKAPEEVTFDYHKASRLGSSAMNIQMLEKLGIPQLEGQKDMITWILQMF 988
Db 721 EEKQQLQKAPEEVTFDYHKASRLGSSAMNIQMLEKLGIPQLEGQKDMITWILQMF 780
QY 989 TVKRLREKSLIVATTSGSESINKDEITHDK 1019
Db 781 TVKRLREKSLIVATTSGSESINKDEITHDK 811

RESULT 5
AAW78576
ID AAW78576 standard; Protein; 1400 AA.
XX
AC
XX
AC
AAW78576;
DT 06-NOV-2001 (first entry)
XX
DE
DE
XX
Human protein SEQ ID NO 1238.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation.
XX
OS Homo sapiens.
XX
XX WQ200157190-A2.
PN
XX
XX 09-AUG-2001.
XX
XX
```

```
PF 05-FEB-2001; 2001WO-US04098.
XX
PR 03-FEB-2000; 2000US-0496914.
PR 27-APR-2000; 2000US-0560875.
PR 20-JUN-2000; 2000US-0598075.
PR 19-JUL-2000; 2000US-0620325.
PR 01-SEP-2000; 2000US-0654936.
PR 13-SEP-2000; 2000US-0663561.
PR 20-OCT-2000; 2000US-0693325.
PR 30-NOV-2000; 2000US-0728422.
XX
XX (HYSE-) HYSEQ INC.
PA
XX
XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX
XX WPI; 2001-476283/51.
DR N-PSDB; AAK51709.
XX
XX Nucleic acids encoding polypeptides with cytokine-like activities,
PT useful in diagnosis and gene therapy -
XX
XX Claim 20; Page 3496-3499; 6221pp; English.
PS
XX
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAM80020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.
XX
XX Sequence 1400 AA;
SQ
XX
XX Query Match 36.7%; Score 3359; DB 22; Length 1400;
XX Best Local Similarity 46.4%; Pred. No. 5.8e-276;
XX Matches 659; Conservative 271; Mismatches 425; Indels 64; Gaps 22;
XX
QY 361 MLWDAKLYKGARKILHELIFSSFFMEMEYKFLFAEFYKVKLOKEVISDDHRSISIT 420
Db 1 MLSDSKLWKGARSYTHQLFMSLLMDLAKYKLLFAVRFAKNTYERLQSDVYDDHRSFVA 60
QY 421 ALSVQMETVPTPLARHLIEEQNVISITETLLEVLPEYLDNRNKNFQGYSDKLG 477
Db 61 DLSVQIFVTPSLARMLITEENLMSIIKTMDHL-RHRDAQGRFQFERYTALQAFKFRV 119
QY 478 YAVTCDLKYILISKPTTWTERLRNQFLEGFRSFLKILTCMQGMEIEIRROVQGHIEVDPDW 537
Db 120 QSLILDLYLVLSKPTWSEDLRQKLEGFDAFLLEKCMQGMDDPITRQVQGHIEPEW 179
QY 538 EAAATAIQMLKNILLMFQEWACACDEELLVAKECHKAVMRCSTSFSSSKTVVQGS-CGH 596
Db 180 EAAFTLQMKLTHVISMMDQWCAKDKVLITAEYAKCLAVLMQCHGYTQGEQITLISICGH 239
QY 597 SLETKSVRVSEDLVSIHLPLSRTLAGLHVRLSRLGAVSRLEHFEVSFDFQVEVLVEYPLR 656
Db 240 SVEITRYCVSQEKYSIHLPLVSRLLAGLHVLSKSEVAYKFPPELLPLSELSPMLIEHLPLR 299
QY 657 CLVLVAQVVAEMWRRNGLSLISQVFFYQDVKCRREMYDKDIIMLIQIGASLMDPNKFLLLV 716
Db 300 CLVLCAQVHAGMWRNGLSLVNIQYIYHNKCRREMYDKDVVLMQIGVSMDDPNHFLMIM 359
QY 717 LQRELAFAFN-----KTISTK--DODLTKQYNTLIEEMQLVLIYIVGERYVPGVGNVTK 769
Db 360 LSRFELYQIFSTPDYGRKFSSEITHKDVQVQNNLTIEEMLYLIIMLGRERFSGVGQVNA 419
```

QY	770	EEVTRRIIHLICIEPMPHSAIAKNLPENENNETHGVNINKVATFKKPGVSGHGVYELK	829
Db	420	TDEIKREIIHOLSIKPMHSELVKSPEDENKETGMSVTEAVAHFKPKLGTGRGVYELK	479
QY	830	DESLKDFNMFYHYKSTQHSAHMOKKRKQENKDEALPPPPPPPCPAFASKVINLLNC	889
Db	480	PECAKEENLYFYHFSRAQBSKAEEAQRKLKQRNREDTALPPVLPFCPLFASLVNILOS	539
QY	890	DIMWYILRTVFERAIDTSDNLWTEGMLQAMAFHIALGLLEEKQOLOKAPEEE-VTFDFYH	948
Db	540	DVMLCINGTTLQWAVEHNGYAWSESMLQRLVHLHIGMALQEQHLENTVEHVVTFTTQ	599
QY	949	KASRLGSSAMN---IQWLLEKLGKIGIQLECGQKDMITWILQFDTVTVRLREKS-CLIVATT	1004
Db	600	KISKPGEAPKNSPSILAMLETQNAPLYEVHKDMIRWLKTFNAVKMKRESSPTSPAET	659
QY	1005	SGSESIKNDITHDKKAERKRAEAAARLHRKIMQAQMSLOKNEFTETHKLMDYNTSEMP	1064
Db	660	EGTIM---EESRDQKAERKRAEATLARLRREKIMQAQMSQMQRHFDENKELFQQTLELD	716
QY	1065	GKEDSIMEESTPAVSYSRIALGPKRGPSVTEKEVLTCTLCQEEQEVKIENNAVLASAC	1124
Db	717	ASTSAVLDH--SPVASDMLTALGPAQTOVPEQRQFVTCILCQEEQEVKVESRAWLAAF	774
QY	1125	VQKSTALTQHRGKPIELSGEALDPLFMDPDLAYGTYTGSGGHVMAVCWQKYEAVOLSS	1184
Db	775	VQRSTVLSKNRSKFIO-DPEKYDPLFMHPDLSCGTHTSSCGHILHACHWQRYEDFSVQAKE	833
QY	1185	QQ-----RTHVDLDFLESGEYGLCPCKSLCNTVPIPILOPOKINSENADALAQLTLAR	1239
Db	834	QRQQRRLRHTS-IDYVNEGFLCPCEJUSNTVPIPL-LPPNRIFNNRLN-FSDQPNLQT	890
QY	1240	WIQTVLARISGVNIRHAKENPIPIEFNCGMDSTLEFHSILSFGVSESSIKYSNSIKEMV	1299
Db	891	WIRTIQQIKALQFLRKEESTPNNASTKNSENVDELQLPGEFRDPRPKPIPSESIKEML	950
QY	1300	ILPATTIYRIGLVKVPDERDPRVPMLTWTSCAFTIOAIENLLGDEGKPLFGALQNQHNG	1359
Db	951	TTFGTATYVGLKVHNEEDPRVPICMWGSCAVTIQSIERILSDKPLFGPLPCRLDDC	1010
QY	1360	LKALMQFAVAORITCPOVLIOKHLVRLLSVVLPNKSEDTPCLLSIDLPHVLYGAVLAPP	1419
Db	1011	LRSLTRFAAAHTVAVSVVQGHCFKPFASLVNPDSHEELPCILDIMFHLVGLVLAEP	1070
QY	1420	SLYWDpDpVLQPSVSSYNHLYLFHLITMAHMLQILL--TVDTGLPLAQVQEDSEEAH	1476
Db	1071	ALOCQD----FSGISLGTGDLHFHVLTWMAHIIQILLTSCTEENGMD--QENPPCEEES	1123
QY	1477	SASSFPAEISQXTSGIGCDIP-GWYLDWLSLKNGITPYLRCAALFFHYLLIGVTPPBELHT	1535
Db	1124	AVLALYKTLHQYT-GSALKEIPSGWLHRSVRAGIMPLFKCSALFFHYLNGVSPDPDIOV	1182
QY	1536	NSAEGYSALCYSLSPTNLFILFOEYWDTVRPLQLORCADPALLMCLQKNTVRYPRK	1595
Db	1183	-PGTSHFEHLKYSLSLPNNLICFLQFNSETMNSLIESWCNRNSVKVYLEGERDAIRYPRE	1241
QY	1596	RNSLIELPDYSCLLNAQSHFRCPRSADDERHPVLCFLCGATILCSNITCCQEIIVGEEV	1655
Db	1242	SNKLINLPEDYSSLLINQASNFSPCKSGGDKSRAPTLCLVCGSLLSQSVCCTQELGEDV	1301
QY	1656	CACIFHALHC-----KARGCAYPAPYLDYEYCTDPLGKRGNPPLHL	1695
Db	1302	GACTAHTYSCGSGVGIFLRVRECQVLFLAGTKGCFYSPYLDYGETOGLRGNPLHL	1361
QY	1696	SRERYKRLHVMWOHCIIIEIARSQETNQMLFGFNWOLL	1734
Db	1362	CKERFKIKQLWHSVTEEIGHAQEANQTLVGIDWQHL	1400

RESULT 6  
 ID AAM79560  
 standard; Protein: 1400 AA.

RESULT 6  
AAM79560  
ID AAM

[illegible]







PT New polypeptide-human ubiquitin relative protein 46.64 for  
 PT treating malignant tumours, inflammations, immunological diseases,  
 PT haemopathy and human immunodeficiency virus infection -  
 XX  
 XX  
 PS Claim 1; Page 25-26 (disclosure); 33pp; Chinese.  
 XX  
 CC The present invention discloses a new kind of polypeptide, human  
 CC ubiquitin relative protein 46.64, polynucleotides for encoding this  
 CC polypeptide, and a DNA recombination process to produce the polypeptide.  
 CC The present invention also discloses the method of applying the  
 CC polypeptide in treating various diseases, such as malignant tumours,  
 CC inflammations, immunological diseases, haemopathy and human  
 CC immunodeficiency virus (HIV) infection. The current sequence represents  
 CC the human ubiquitin relative protein 46.64 of the invention.  
 XX  
 XX Sequence 424 AA;  
 SQ

Query Match 23.7%; Score 2168; DB 23; Length 424;  
 Best Local Similarity 98.8%; Pred. No. 2.2e-175;  
 Matches 417; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 501 MFLGFRSFLKILTCMQGMEIRROVGHIEVDPDWEAAIAIQMLKNILLMFQEWCA 560  
 Db  
 1 MFLGFRSFLKILTCMQGMEIRROVGHIEVDPDWEAAIAIQMLKNILLMFQEWCA 60

QY 561 DEILLVAYKECHKAVMRCSTSFISSTKVQSCGHSLETKSYRVEDLVSIHLPLSRTL 620  
 Db  
 61 DEILLVAYKECHKAVMRCSTSFISSTKVQSCGHSLETKSYRVEDLVSIHLPLSRTL 120

QY 621 AGLHRLSLGLGAVSRHLEFVSFEDFQVEVLVEPLRCLVLVAQVVAEMRRNGLSLISQV 680  
 Db  
 121 AGLHRLSLGLGAVSRHLEFVSFEDFQVEVLVEPLRCLVLVAQVVAEMRRNGLSLISQV 180

QY 681 FYYQDVKCREMYDKDIIMLQIGASLMDPNKFLLLVLQRYELAEAFNKTIISTKDQDLIK 740  
 Db  
 181 FYYQDVKCREMYDKDIIMLQIGASLMDPNKFLLLVLQRYELAEAFNKTIISTKDQDLIK 240

QY 741 YNTLIEMLQVLIYIGERYVPGVGNVTKKEVTMREIIHLICIEPMPHSAIAKNLPENEN 800  
 Db  
 241 YNTLIEMLQVLIYIGERYVPGVGNVTKKEVTMREIIHLICIEPMPHSAIAKNLPENEN 300

QY 801 NETGLENVINKVATFKKPGVSGHGYVELKDESLKDFNMVYHYSKTQHSKAEHMQKRRK 860  
 Db  
 301 NETGLENVINKVATFKKPGVSGHGYVELKDESLKDFNMVYHYSKTQHSKAEHMQKRRK 360

QY 861 QNKDEALPPPPPPPCPAFSKVINLLNCDIMMYILRTVFERAIDTDSNLWTGMLQMAF 920  
 Db  
 361 QNKDEALPPPPPPPCPAFSKVINLLNCDIMMYILRTVFERAIDTDSNLWTGMLQMDW 420

QY 921 HI 922  
 Db  
 421 EV 422

RESULT 10  
 AA084353  
 ID AA084353 standard; Protein; 333 AA.  
 XX  
 AC AA084353;  
 XX  
 DT 25-MAR-1999 (first entry)  
 XX  
 DE Partial human ubiquitin-protein ligase, Ubr1.  
 XX  
 KW Ubiquitin-protein ligase; Ubr1; human; ubiquitylation; degradation;  
 KW N-end rule pathway; stress-related muscle wasting; inhibitor; screen.  
 XX  
 XX Homo sapiens.  
 OS  
 XX US5861312-A.  
 PN  
 XX 19-JAN-1999.  
 PD  
 XX

PF 02-DEC-1997; 97US-0982956.  
 XX  
 PR 02-DEC-1997; 97US-0982956.  
 XX  
 XX (CALY ) CALIFORNIA INST OF TECHNOLOGY.  
 PA  
 XX Kwon YT, Varshavsky A;  
 PI  
 XX WPI; 1999-130395/11.  
 DR N-PSDB; AAX03300.  
 DR  
 XX Mouse and human Ubr1 cDNA - useful for producing recombinant Ubr1  
 PT polypeptides  
 PT  
 XX Disclosure; Columns 27-30; 18pp; English.  
 PS  
 XX The present sequence represents a partial ubiquitin-protein ligase called  
 CC Ubr1. The Ubr1 enzymes are involved in protein ubiquitylation and  
 CC ultimate degradation through the N-end rule pathway and have been  
 CC linked to stress-related muscle wasting. Recombinant Ubr1 polypeptides  
 CC can be used to screen for inhibitors of muscle wasting when this is  
 CC associated with the N-end rule pathway.  
 XX  
 SQ Sequence 333 AA;

Query Match 18.8%; Score 1716; DB 20; Length 333;  
 Best Local Similarity 99.7%; Pred. No. 4.8e-137;  
 Matches 332; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 707 MDPNKFLLLVQRYELAEAFNKTIISTKDQDLIKQYNTLIEMLQVLIYIGERYVPGVGN 766  
 Db  
 1 MDPNKFLLLVQRYELAEAFNKTIISTKDQDLIKQYNTLIEMLQVLIYIGERYVPGVGN 60

QY 767 VTKEVTMREIIHLICIEPMPHSAIAKNLPENENNETGLENVINKVATFKKPGVSGHGY 826  
 Db  
 61 VTKEVTMREIIHLICIEPMPHSAIAKNLPENENNETGLENVINKVATFKKPGVSGHGY 120

QY 827 ELKDESLKDFNMVYHYSKTQHSKAEHMQKRRKQENKDEALPPPPPPPCPAFSKVINL 886  
 Db  
 121 ELKDESLKDFNMVYHYSKTQHSKAEHMQKRRKQENKDEALPPPPPPPCPAFSKVINL 180

QY 887 LNCDDIMMYILRTVFERAIDTDSNLWTGMLQMAFHILALGLLEEKQQLQKAPEEVTFDF 946  
 Db  
 181 LNCDDIMMYILRTVFERAIDTDSNLWTGMLQMAFHILALGLLEEKQQLQKAPEEVTFDF 240

QY 947 YHKASRLGSSAMNIQMLLEKLGIPQLGQKDMITWILQMDTVTKRLREKSLIVATTSS 1006  
 Db  
 241 YHKASRLGSSAMNIQMLLEKLGIPQLGQKDMITWILQMDTVTKRLREKSLIVATTSS 300

QY 1007 SESIKNDEITHDKEKAERKRAEAAARLHRQKIM 1039  
 Db  
 301 SESIKNDEITHDKEKAERKRAEAAARLHRQKIM 333

RESULT 11  
 AAB31163  
 ID AAB31163 standard; Protein; 333 AA.  
 XX  
 AC AAB31163;  
 XX  
 DT 02-APR-2001 (first entry)  
 XX  
 DE Amino acid sequence of a partial human Ubr1 protein.  
 XX  
 KW Ubr1; E3-type protein; ubiquitin system; ubiquitin-protein ligase;  
 KW N-end rule pathway; intracellular pathogen; Lysteria monocytogenes;  
 KW Yersinia enterocolitica; muscle wasting; infection.  
 XX  
 OS Homo sapiens.  
 XX  
 PN US6159732-A.  
 XX  
 PD 12-DEC-2000.

XX 11-JAN-1999; 99US-0228317.  
PF  
XX  
PR 02-DEC-1997; 97US-0982956.  
XX  
XX (CALY ) CALIFORNIA INST OF TECHNOLOGY.  
PA  
XX Kwon YT, Varshavsky A;  
PI  
XX WPI; 2001-090278/10.  
DR  
XX N-PSDB; AAC86934.  
DR  
XX  
PT Inhibiting the N-end rule pathway in mammalian cells for treating  
PT infections and various diseases associated with muscle tissue wasting,  
PT by inhibiting the expression of Ub1 gene  
XX  
XX Example; Column 27-30; 18pp; English.  
PS  
XX  
XX The present sequence represents a partial Ub1 enzyme. Ub1 is an E3-type  
CC protein of the ubiquitin system. Specifically, it is a ubiquitin-protein  
CC ligase. The enzyme is specific for destabilising residues exposed at  
CC the N-terminus of protein substrates. Inhibition of the expression of  
CC Ub1 gene in a cell results in inhibition of the N-end rule pathway.  
CC The method is used for treatment of mammalian cells infected with an  
CC intracellular pathogen, e.g. Lysteria monocytogenes or Yersinia  
CC enterocolitica. Inhibition of N-end rule pathway is also useful for  
CC treating various diseases associated with wasting of muscle tissue and  
CC infections.  
XX  
XX Sequence 333 AA;  
SQ  
Query Match 18.8%; Score 1716; DB 22; Length 333;  
Best Local Similarity 99.7%; Pred. No. 4.8e-137;  
Matches 332; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
Qy 707 MDPNKFLLVLRQYELAEAFNKTISTKDDLIKQYNTLIEMLQVLIYIGERYVPGVN 766  
Db 1 MDPNKFLLVLRQYELAEAFNKTISTKDDLIKQYNTLIEMLQVLIYIGERYVPGVN 60  
Qy 767 VTREVTMRRIHLLCPEMPHSAIAKNLPENNETGLENVINKVATFKPGVSGHGV 826  
Db 61 VTREVTMRRIHLLCPEMPHSAIAKNLPENNETGLENVINKVATFKPGVSGHGV 120  
Qy 827 ELKDESLEKDFNMYHYHYSKTOHSAEHMQKRRKQENKDEALPPPPPEFCPAFSKVINL 886  
Db 121 ELKDESLEKDFNMYHYHYSKTOHSAEHMQKRRKQENKDEALPPPPPEFCPAFSKVINL 180  
Qy 887 LNCDDIMMYILRTVPERAIDTDSNLWTEGMLQMAFHIALGLLEKQQLQKAPEEVTFDF 946  
Db 181 LNCDDIMMYILRTVPERAIDTDSNLWTEGMLQMAFHIALGLLEKQQLQKAPEEVTFDF 240  
Qy 947 YHKASRLGSSAMNTOMLEKLGIPQLEGQKDMITWILQMDFTVKRLREKSLIVATTSG 1006  
Db 241 YHKASRLGSSAMNTOMLEKLGIPQLEGQKDMITWILQMDFTVKRLREKSLIVATTSG 300  
Qy 1007 SESIKNDEITHDKEAERKKAERAAARLHRQKIM 1039  
Db 301 SESIKNDEITHDKEAERKKAERAAARLHRQKIM 333  
RESULT 12  
ABB90168  
ID ABB90168 standard; Protein; 452 AA.  
XX  
XX ABB90168;  
AC  
XX  
XX 24-MAY-2002 (first entry)  
DT  
XX  
XX Human polypeptide SEQ ID NO 2544.  
DE  
XX  
XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;  
KW antiinflammatory; hepatotropic; antidiabetic; antiinflammatory; antitumor;  
KW vulnery; anticonvulsant; antibacterial; antifungal; antiparasitic;

KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;  
KW neurological disease; infection; human; secreted protein.  
XX  
OS Homo sapiens.  
XX  
PN WO200190304-A2.  
XX  
PD 29-NOV-2001.  
XX  
XX 18-MAY-2001; 2001WO-US16450.  
PF  
XX 19-MAY-2000; 2000US-205515P.  
PR  
XX (HUMA-) HUMAN GENOME SCI INC.  
PA  
XX Birse CE, Rosen CA;  
PI  
XX WPI; 2002-122018/16.  
DR  
XX N-PSDB; ABL90577.  
DR  
XX Novel 1405 isolated polypeptides, useful for diagnosis, treatment and  
PT prevention of neural, immune system, muscular, reproductive,  
PT gastrointestinal, pulmonary, cardiovascular, renal and proliferative  
PT disorders -  
XX  
XX Claim 11; SEQ ID NO 2544; 208lpp + Sequence Listing; English.  
PS  
XX The invention relates to novel genes (ABL89449-ABL90853) and proteins  
CC (ABB9040-ABB9044) useful for preventing, treating or ameliorating  
CC medical conditions e.g. by protein or gene therapy. The genes are  
CC isolated from a range of human tissues disclosed in the specification.  
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful  
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast  
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone  
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;  
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune  
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's  
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;  
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and  
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
CC and parasitic infections.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 452 AA;  
SQ  
Query Match 11.7%; Score 1071.5; DB 23; Length 452;  
Best Local Similarity 44.9%; Pred. No. 6.4e-82;  
Matches 207; Conservative 79; Mismatches 142; Indels 33; Gaps 7;  
Qy 1298 MVILFATTYIRGLKVPDPDRPRVPMLTWSTCAFTIOAENLLGDEKPLFGALQNRH 1357  
Db 1 MLTFETATYKVGKLVHPNEEDPRVPMCMGSCAYTQSTERLSDKPLFGPLCRLD 60  
Qy 1358 NGIKALMOFAVAGRITCPQVLQKHVLRLLSVVLNPKISDPTCLLSIDLHVLVGA 1417  
Db 61 DCLSLRTFRAAAHWTVASVSVQGGCKFLASLPNDSDHEELPCILIDHFLVLGLVLA 120  
Qy 1418 FPSLYWDDPVDLQPSVSSSYNHLFLHITMAHMLQILL---TVDTGLPLAQVQDESEE 1474  
Db 121 FPALQCD-----FSGISLGTGDLHIFHLVTHMAHIIQLLTSCTEENGMD--QENPPCEE 173  
Qy 1475 AHSASSFFAEISQYTSGSGCCDIP-CWYLMVSLKNGITPYLRCAALFFHLLGVPPEEL 1533  
Db 174 ESAVLALYKTLHQYT-GSALKETPSGWHLWRSVRAGIMPFKCSALFFHYLNGVSPSPDI 232  
Qy 1534 HTNSABGEVSALCSYLSLPTNLFLLFQGYWDTVRPLLORRCADPALLNCLKQNTVVRYP 1593  
Db 233 QV-PGTSHEHLCSYLSLNNLCLFQENSEINLSWCNRSNVEKRYLEGERDAIRYP 291  
Qy 1594 RKRNSLIELPDDYSCLLNQASHFRCPRSADDERKHPVLCLFCGAILCSQNICQEIYNGE 1653



Db 292 RESNKLINLPEDYSSLIQASNFSCPKSGGDSRAPTCILVCGSLSCSSQYCCQTELEGE 351  
QY 1654 EVGACIFHALHC-----KARGCAYPAPYLDYGETDPGLKRGNPL 1693  
Db 352 DVGACTAHTYSCGSGVGIFLRVRECVQLFLAGTKGCFSPYLDYGETDQGLRRGNPL 411  
QY 1694 HLSRERYKHLVWQOHCIIIEIARSQETNOMLFGFNQOLL 1734  
Db 412 HLCRERFKKIOKLWHOHSVTEIGHAQEANTLVGIDWQHL 452

RESULT 13  
ABG05917  
ID ABG05917 standard; Protein; 487 AA.  
XX  
AC ABG05917;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #5908.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.

XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
XX  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DR WPI; 2001-639362/73.  
DR N-PSDB; AAS70104.

XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX  
PS Claim 20; SEQ ID No 36276; 103pp; English.  
XX

CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 487 AA;

Query Match 9.6%; Score 877; DB 22; Length 487;  
Best Local Similarity 59.2%; Pred. No. 2.7e-65;  
Matches 189; Conservative 0; Mismatches 0; Indels 130; Gaps 2;  
QY 643 EDFQVEVLVEYPLRCLVLVAQVVAEMRRNGLSLSISQVFFYQDYKCREMYDKDIIML-- 700  
Db 3 EDFQVEVLVEYPLRCLVLVAQVVAEMRRNGLSLSISQVFFYQDYKCREMYDKDIIMLQK 62  
QY 701 -----QIGASLMDPNKFLLLVQRYELAAEFNKITSTKDQDLIKQYNTLIEMLQVL 752  
Db 63 KYDAGFLSQIGASLMDPNKFLLLVQRYELAAEFNKITSTKDQDLIKQYNTLIEMLQVL 122  
QY 753 IYIVGERYPVGNGVTKEEVTMRRIIHLCTEPMPHSAIAKNLPENNETGLENVINKV 812  
Db 123 IYIVGERYPVGNGVTKEEVTMRRIIHLCTEPMPHSAIAKNLPENNETGLENVINKV 168  
QY 813 ATEKKPCVSGHGVYELKDESILKDFNMYFYHYSKQHSKAEHMQKRRKQENKDEALPPPP 872  
Db 169 ----- 168  
QY 873 PPEFCPAFSKVINLLNCDIMMYILRTVFERAIDTDSNLWTEGMLQMAFHILALGLEEKQ 932  
Db 169 -----AFHILALGLEEKQ 182  
QY 933 QLQKAPEEEVTFDFYHKAS 951  
Db 183 QLQKAPEEEVTFDFYHKAS 201

RESULT 14  
ABG05881  
ID ABG05881 standard; Protein; 258 AA.  
XX  
AC ABG05881;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #5872.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DR WPI; 2001-639362/73.  
DR N-PSDB; AAS70068.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX  
PS Claim 20; SEQ ID No 36240; 103pp; English.  
XX  
CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags

for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic amino acid sequences of the invention.

Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://ftp.wipo.int/pub/published_pct_sequences).

[illegible]

Qy	758	ERYVPGVGVNTKKEVNTMREIHHLLCIEPMPHSAIAKNLPENENNFTGLENVINKVATFKK	817
Db	852	ERRFCGLSTA---ESLRREIIFLATGDETHSOLVKSPLRDLKSDELOEVLDDSVYCN	908
Qy	818	PGVSGHGVYELKDSELSKDFNMVYFYHSKTHQSKAEHMOKRRKQENKDEALPPPPPEFC	877
Db	909	PSGMNKGYSLQSSCWKEDLY-----HPRWQSRDLOSAERF-----SRYC	950
Qy	878	-----PAFSKVINLLNCIDIMVILKTFERAITDTSNLTW---EGMLQW	918
Db	951	GVSALTTQLPWRMIYPLKGLARIGCTKATFOIISALYALQSTSVKSRAPDGVLT	1010
Qy	919	AFHILALGLEEKQOIQKAPEE---EVTDFYHKAS-----RLGSSAMNIQMLEKLKG	969
Db	1011	ALQLLSLSDICTQORQSNQDCCLSENSPILELAGLEITIGTAQTEKESLLSLVSLMK	1070
Qy	970	IPOLEGK-----DMITWI---LQMEDTV---KRLREKSLCIAVATSGSESTKDE	1014
Db	1071	TRMGDRGHOFPEPGSCNISWIGNLLKKFSAYDSVCMNLLQSLAPEVQSGFDKVMGS	1130
Qy	1015	ITHDKGAERKRAEAAARLHRQRIQAOMSALOKNFIETHKLMYDNTSEMPGKE---	DSI 1070
Db	1131	TSDEKRRKAKERQAA-----IMAKMRAEQSKFLSTLSSMDD--DPRSEFTSDSV	1181
Qy	1071	MEESTPANSVDYGRIALGP-KRGP-----SVTEKEVLTC	1103
Db	1182	MEHDSETAREVSGSLCHDPKDPVPSFLIFLOWGSMTDIICDC	1225

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	176	1.9	2662	4	US-09-595-684B-31		Sequence 31, Appl
2	172	1.9	3878	4	US-09-914-259-11		Sequence 11, Appl
3	156.5	1.7	2710	1	US-08-480-604A-6		Sequence 6, Appl
4	156.5	1.7	2710	2	US-08-405-496A-6		Sequence 6, Appl
5	156.5	1.7	2710	3	US-08-915-136-6		Sequence 6, Appl
6	156.5	1.7	2710	4	US-08-957-310-6		Sequence 6, Appl
7	156.5	1.7	2710	4	US-10-011-366-6		Sequence 6, Appl
8	156	1.7	3433	4	US-09-091-501B-10		Sequence 10, Appl
9	150.5	1.6	1388	4	US-09-572-191-2		Sequence 2, Appl
10	150.5	1.6	1388	4	US-09-723-262-2		Sequence 2, Appl
11	150.5	1.6	1388	4	US-09-723-219-2		Sequence 2, Appl
12	147.5	1.6	3248	1	US-08-353-700-1		Sequence 1, Appl
13	147.5	1.6	3248	5	PCR-US95-16216-1		Sequence 1, Appl
14	140.5	1.5	959	4	US-09-914-259-67		Sequence 67, Appl
15	140.5	1.5	1708	1	US-08-493-093-2		Sequence 2, Appl
16	140.5	1.5	1708	1	US-08-508-836A-2		Sequence 2, Appl
17	139.5	1.5	2482	1	US-08-328-254-6		Sequence 6, Appl
18	139	1.5	976	3	US-09-104-324B-4		Sequence 4, Appl
19	137	1.5	2210	4	US-09-309-572-7		Sequence 7, Appl
20	134	1.5	1242	4	US-09-107-532A-5241		Sequence 5241, Appl
21	133.5	1.5	961	1	US-09-914-259-66		Sequence 66, Appl
22	133.5	1.5	3056	1	US-08-508-836A-8		Sequence 8, Appl
23	133.5	1.5	3056	2	US-08-629-001A-3		Sequence 3, Appl
24	133.5	1.5	3056	2	US-08-874-266-2		Sequence 2, Appl
25	133.5	1.5	3056	3	US-08-642-274D-3		Sequence 3, Appl
26	133.5	1.5	3056	3	US-08-952-127-3		Sequence 3, Appl
27	133.5	1.5	3056	3	US-08-952-014C-3		Sequence 3, Appl

```
QY 437 --IEQNVISVITETLLEVLPE--YLDNRNKNFQGYSDQKLGRRYAVICDLKYLISK- 491
Db 971 ESLKHQETINTLKSISEEYRNLMHEENTGETKDEFQKMWG-----IDKQDLEAKN 1024
QY 492 -PIWTERLBMQFLEGRPSFLKILTCMQGMEETRRQVGHIEYDWDWEAAIAIOMQLK-- 548
Db 1025 TOTLTADVKNDEITEEQOR--KIFSLTOEKNELQOML-----ESVIAEKEQLKTD 1071
QY 549 ---NILLMFOECACDEELLVA--YKECHKAYMRCSTSFSSSKTVVQSCG-----HS 597
Db 1072 LKENIENTIEN---QEBLLGDELKQOEIVAEQKHAHKKEGELSRCTCDRLAEVEEK 1127
QY 598 LETKSVRSE---DLVSIHPLS-----RTLGLHVLRLSLGAVSRL 636
Db 1128 LKESQOLQEQOQLLVQEMSEMOKKINEIENLKNELNKKELTLEHMETERLELAQKL 1187
QY 637 HEFVSFEDFOVELVLEPLRCLVLVAOVVAEMRRNGLSLISVYFYQDVKCREMYDKD 696
Db 1188 NE--NYEE-----VKSITKE--RKVLKELQKSFETERDHLRGYIREIE 1236
QY 697 IIMLOIGASLMDPNKFLVLLVQRY-ELAEAFNKTISTK-----DQDLIKQYNTLIEML 749
Db 1227 ATGLQTKEL---KIAHILKHEQETIDELRRSVSEKTAQIINTQDLKSHTKLOEE-- 1280
QY 750 QVLIYIVGVRYPGVGVNVTKEEVTMRIHL-----LCIEPMPSAIAKNLPEN 798
Db 1281 -IPVLEEQLLVNKKVSTQETMNELELLTQSTTKDSTTLARIEMERLRUNEKFOES 1339
QY 799 ENNETGLENIENVKATPKPGVSGHYVELKDESKDFNMVYFH---YKSTQSKAEHM 854
Db 1340 QEEIKSLTKERDNLTKIE-----ALEVKHDOKE-----HIRETLAKIQESQSKQE 1386
QY 855 QKRRKQENKDEALPPPPPEFCOPAFSKVINLNCDIMMYL-----RTV----- 899
Db 1387 QSLNMKEKDNETTKIVSEMEQFKPKOSA---LLRIETIEMGLSKRLQESHDEMSVAKE 1443
QY 900 -----FRADTDSNLWTEGMLQ-MAPHILALGLLEKEQLOKA-----PEEEVTFDFYH 948
Db 1444 DDLQRLQEVLOQESDQKENIKEIVAKH-----LETEELKVACHCLKEQEEINELRV 1497
QY 949 KASRLGSSAMNIOMLLE-----KLKGIPQLEGQDMITWILQDMFTVKRLREKSLI 1000
Db 1498 NLSEKETEISTIQOLBAINDKLNKIQEYIEKEEQLN-IQISEVQENVNELQKPEHR 1556
QY 1001 VATTGSESINKD--EITHD-----KEKAERKRAEAAARLHR-----OKIM 1039
Db 1557 KAKDSALQSTESKMLELTNRLOESQEBIQIMIKEEMKRVQEOALQIERDQKENTREIV 1616
QY 1040 AQMSALQ-----KNFIETHKL-----MYDN 1059
Db 1617 AKMKEQKEYQFLKMTAVANETOEKMEIEHLKEQFETQKLNLENIETENIRLQIILHEN 1676
QY 1060 TSEMPG---KEDSIMPEESTPAVSDYSRIALGPKRGPSTVEKVLVCILOEQEYVKI- 1114
Db 1677 LEEMRSTVKRDDRDSRVEETLKVERDQKCNL-----RETITRDL-EKQEBELKIV 1725
QY 1115 -----ENNAMY--LSACVQKSTALTOHRGKPIELSGEAL 1146
Db 1726 HMHLKHEQETIDKLRGIVSEKTEINISNMQDLHNSDAL 1764
```

## RESULT 2

```
US-09-914-259-11
; Sequence 11, Application US/09914259
; Patent No. 6495336
; GENERAL INFORMATION:
; APPLICANT: Makowski, Lee
; APPLICANT: Hyman, Paul
; APPLICANT: Williams, Mark
; TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES
; FILE REFERENCE: 8471-010-999
; CURRENT APPLICATION NUMBER: US/09/914,259
; CURRENT FILING DATE: 2000-11-21
```

```
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 3878
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-914-259-11

Query Match 1.9%; Score 172; DB 4; Length 3878;
Best Local Similarity 19.3%; Pred. No. 4.7e-05;
Matches 249; Conservative 188; Mismatches 409; Indels 446; Gaps 61;

QY 187 LNEEVIVCARKIFPSVIKVVVEMTWEEKEKLPPE-LOIREKNERVYCVLFNDEHHSYDH 245
Db 908 LNEELHLQ--RINPTTVK--MKSSVFDDEKTFVATLEMGVEVVEKOTTELMKEKLEVTKRE 963
QY 246 VIYSLQALDCELABAQLH--TTAIDKEGR--RAYKAGAYAACQAKEDIKSH--SENVYS 299
Db 964 KLELSQRLSDLSEQLKQKHGEISPLNEEVKSLQKEQVSLRCREL-EIINHNRANVQ 1022
QY 300 QHPLHVEVL-----HSEIMAHOKFALRLGSMWNKIMSYS 334
Db 1023 SCDTQVSSLLDGVVTMTSRGAEGSVSKVKNKSFGEESKIMVEDKV-----SPENMTVGEES 1077
QY 335 DFRQIFC-----QACLRPEPDSNPCLISRLMLW-----DAKLYKGARKILHELIES 381
Db 1078 KOEQILDLHPLSVTKESSLRATQPSNDKLOKELNVLKEQNDLRLQMEAQRICLSLVYS 1137
QY 382 SF-----FMEMEYKK-----LFAMEFVKYKQLOKEYSIDHDHRSISITALSVO- 425
Db 1138 THVDQVREYMEKDKALCSLKEELIFAOE-EKIKELQKIH-----QLELQTMKTOE 1188
QY 426 -----MFTVPTLARIHLIEEQNVISVITETLLEVLPEYL-----DRNNKFN 466
Db 1189 TGDEGKPLHLIGLKQKAVSEE--CSYFLOPLCSVLGEYTPALCKEYNAEDKENSQDY 1245
QY 467 QGYSDKLGRRYAV-ICDLK---YILISKPT-----IW---TERLRMQFL 504
Db 1246 ISENEPELDQRYREVQDFQENMHTLLNKVTEYKNLLVLQTRLSKQQTGDGKLEFG 1305
QY 505 E-----GFRSFLKIITCMQGMEEIRRVQGHIEVDPDWEAAIAIOMKLNILLMFQEMCA 559
Db 1306 EENLPKETEFLSIHQMTNLEDI-----DVNHKSKLSSLODLKTEKLEEGV--- 1352
QY 560 CDEELLVAYKECHKAVMRCSTSFSSSKTVVQSCG-----HSLKTSYRVSSEDLVSI 612
Db 1353 --QEL-----ESLISSLQOQLKETEYAEIHCQLKRLQAVSESTVPP 1394
QY 613 HLPL-----SRTL-AGLHVRLSRLGAVSRHLHEFVSFEDFOVEVLVE--YPLRCLV 659
Db 1395 SLPVDSVVITESDAQRTMYPGSCVRKKNIDGTTFESGEGFVKETNIVKLEKQYQBLEE 1454
QY 660 LVAOVVAEM---WRRNGLSLIS-----OV 680
Db 1455 EVAKVIVMSIAFAQQOTELSRISGGKENTASSQAHAVCQEQHIFNEMKLSODQIGFQT 1514
QY 681 FYYQDVKCR-----EEMYDKDIIML--- 700
Db 1515 FETVDVKFEKPKPLSKELGEGHKEILLNSDPHDIPESKDCVLTISEEMFSKDKTFIVR 1574
QY 701 -----QIGASLMDPNKFLVLLVQRYELAEAFNKTISTKQDOLIKQY----- 741
Db 1575 QSIHDEISVSSMDASRLMLNNEQLE-----DMRQELVROVQEQHQQAQORSIDN 1624
QY 742 NTLIEEMLOVLIYIVGVRYPGVGNVTKEEVTMRIHL-----LCIEPMPSAIAKNLP 796
Db 1625 ENLVSESEREVLL-----BELEALKOLSLAGREKJCCELRNSTQQTQNGN 1668
QY 797 ENE-----NNETGLENIENVK---ATFFKPGVS 821
Db 1669 ENQGEVEEOTFKKEKELDRKPEDVPPPEILSNRYALOKANNRLKLLILEVVKTAAVEETI 1728
QY 822 GHGVVELKDESUKDFNMVYHYSKTQ-----HKAEMHMQKRRKQEN---KDEALP--- 869
```

Db 1729 GRVLGILDRS-----SKSQSSASLIWRSEAEASVSKVHEEHTRVTDPSY 1777  
QY 870 --PPPPPCPAFSKV-----INLNCIDIMYI---LRTVFERAI 904  
Db 1778 SGSDMPRNDINMWSKVTEEGTSLQRLVRSFGAGTEIDPENEELMLNISRLQAAYEKLL 1837  
QY 905 DFDLSLWTEGMLQMAFHILALGLEEK-QOLQKAPK-----EETFDYHFKASRLGSSAMN 959  
Db 1838 EASE--TSSQLEHA-KVTQTELMRESFRQKEATESLKQCDELRLHEER---AREQ 1891  
QY 960 IOMLEKLGIPQLEGQKDMIT-----WILQMFDTVKRLREKSLIVATTSGSP-SIKNDE 1014  
Db 1892 LAVELSKAEGV--IDGADKTLFERQIQEKTIDIDRLQE--LLCASNRLQELAEQQQ 1947  
QY 1015 ITHDKEKAERK--KAEARLHROKIMQMSALOKNFIE---THKMYDNTSEMPGKE 1067  
Db 1948 IQEERELLSROKAMKAEGVP-EQQLQETELKMEKLEVOQCAEKVRDDLQKQVKALE 2006  
QY 1068 DSI-----MEESTPAVSDYSRIALGPKRGPSVTEKEVLTCILCOEEQEVKIENNA 1118  
Db 2007 IDVEQVSRFIELEQKNTFMDL-----RQOQNALEKQLEKMRKFLDEQAIDREHER 2059  
QY 1119 MYLSACVOK-----STALTQHRKPIE 1140  
Db 2060 DVFOEQIEQKLEQQLKVVPFQFISEHQTREVE 2091  
RESULT 3  
US-08-480-604A-6  
; Sequence 6, Application US/08480604A  
; Patent No. 5736139  
; GENERAL INFORMATION:  
; APPLICANT: KINK, JOHN A.  
; APPLICANT: THALLEY, BRUCE S.  
; APPLICANT: PADHYE, NISHA V.  
; APPLICANT: FIRCA, JOSEPH R.  
; APPLICANT: STAFFORD, DOUGLAS C.  
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND  
; TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE  
; NUMBER OF SEQUENCES: 32  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MEDLEN & CARROLL, LLP  
; STREET: 220 MONTGOMERY STREET, SUITE 2200  
; CITY: SAN FRANCISCO  
; STATE: CALIFORNIA  
; COUNTRY: UNITED STATES OF AMERICA  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/480,604A  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/422,711  
; FILING DATE: 14-APR-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/405,496  
; FILING DATE: 16-MAR-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/329,154  
; FILING DATE: 25-OCT-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/161,907  
; FILING DATE: 02-DEC-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/985,321  
; FILING DATE: 04-DEC-1992  
; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/429,791  
; FILING DATE: 31-OCT-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: INGOLIA, DIANE E.  
; REGISTRATION NUMBER: 40,027  
; REFERENCE/DOCKET NUMBER: OPHD-01763  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 705-8410  
; TELEFAX: (415) 397-8338  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2710 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-480-604A-6  
Query Match 1.7%, Score 156.5; DB 1; Length 2710;  
Best Local Similarity 18.6%; Pred. No. 0.00065;  
Matches 189; Conservative 154; Mismatches 354; Indels 317; Gaps 44;  
QY 288 KEDIKSHENVSOHPLHVEV---LHSEIMAHQKFAIRLGSMMKIMSYSSDFRQIFCOAC 344  
Db 355 KSEIFSKLENLVNVDLEIKIAFALGSVI--NOALISKOGSYLTNLVIEQVKNRYQFLAQH 412  
QY 345 LREPDSENCPLISRLMLDAKLYKGARKILHELIFSSFFMEMEYKFLFAMEFYKYYKQL 404  
Db 413 LNPAIESDNN-----PTDTTKTFHDSLFNSATAE---NSMFLTKIAPY---L 453  
QY 405 QKEYISDDHRSISITA-----LSVQMTV-PTL-ARHLIE---BONVISVITE 448  
Db 454 QVGFMPERS-TISLSPGAYASAYDFINLOENTIEKTLKASDLIEKFFENNLSQLTE 512  
QY 449 TLLEVLPEYLDNRNKNFQYSDQKLGRVYA-----VICDLKYLISK-PTIWTE 497  
Db 513 QEINSLWSPQASAKYQFEKYVVDYTGGSLSSENGVDNKNKNTALDKNYLLNKNIPSNVVE 572  
QY 498 RLMOQFLEGRSFLKILTCMOGMBEIRROVGOHVEDPDWEAAIAIOMLKNILLMFOEW 557  
Db 573 E-----AGSKNYVHYIQLQGGDISYEATCNLFKNP--KNSIIQRNMNESAKSY--F 622  
QY 558 CACDEELLVAYKECHKAVMRCSTSFISSSKTVVSCGH-----SLETKSYR 604  
Db 623 LSDDGESIL-----ELNK--YRIPRLKNKVKVTFIGHGKDEPNTSEFARLSVDSLNE 676  
QY 605 VSEDLVSIHLPLSRTLGLVRLSRLGAVSRHLFEVDFEDQVEVLVYELRCLVLAQV 664  
Db 677 ISSFLDTIKLDISPK---NVEVNLGC-----NMFYSY-DFNVE--ETYPCKLLLSIMDK 724  
QY 665 VA-----EMWRNGLSLISQVFFYQDVVKCREMY--- 693  
Db 725 ITSTLPDVNKNSTITIGANQVEVRINSEGRKELLASHGKWINKEAIMSDLSKEYIFPDS 784  
QY 694 -----DKDIIMLQICASLMDNPKFLLVLQ-----RYELAE 724  
Db 785 IDNKLKAKSNIPCLASISEDIKTLDDASVSPOTKFIILNKLKNISSIGDYIYIEKLE 844  
QY 725 AFNKTISTKDDLIKOYNTLIEEMQLVLIYIVGERYPVGVNVTKEEVTREIHLCLIE 784  
Db 845 PVKNIHNSIDDLIDEFNL-----ENVSELYELAKL----- 877  
QY 785 PMPHSAIAKNLPE-----NENNETGLENVINKVATFKKPGVSGHGVYELKDESUKD 835  
Db 878 -----NNLDEKYLISFEDISKNNSTYSVRINK-----SNGESVYVETEKEI-- 919  
QY 836 FNMVYFHYSK-----TQHSKAEIMQKRRKQENKDEALPPPPPFCEPAFSKV 883  
Db 920 FSKYSEHITKEISTIKNSIITVDNGLNDLQDLHTSQ----- 957  
QY 884 INLLNCIDIMYIILTVFERAIDTD--SNLWTEGMLQMAFHILALGLEEKQOLQKAPKEE 941  
Db 958 VNTLNA---AFFIQSLIDYSSNKNVDLNDLSTSVKQVLAQFLSTGLNTIYDSIQLV----- 1010

QY 942 VTDFYHKASRLGSSAMN--IOMLEKLEKIGIPQLEGOKDMITW---ILOMFDTVKRLREK 996  
Db 1011 -----NLISNAVNDTINVLPITTEGIPVSTILDGNGLGAARKELLDHDPDLKK 1060  
QY 997 -----SCLIVATTSGSESINKNDEITHDKEAERKKAERKAEARLHROKIMQMS 1043  
Db 1061 ELEAKVGVLAINNSLSIAATVASIVGI-GAEVT-----IFLLPIAGISAGIP 1106  
QY 1044 ALQKNFIETHKLMYDNTSEMPGKEDSIMEEESTPAVSDYSRIALGPKRGPSVTEKE-VLT 1102  
Db 1107 SLVNN-----ELILHDKATSVVNYFNHLSSESKYGLPKTDDDKILV 1147  
QY 1103 CILCOEQEVKIENNANVLSCVQKSTALTQHRKPIELSGEALDPLFMDPDLA 1156  
Db 1148 PIDDLVISEIDFNNSIKLGTG-----NILAMEGGSGHTVTGN-IDHFFSSPSIS 1196

## RESULT 4

US-08-405-496A-6  
; Sequence 6, Application US/08405496A  
; Patent No. 5919665  
; GENERAL INFORMATION:  
; APPLICANT: WILLIAMS, JAMES A.  
; TITLE OF INVENTION: VACCINE FOR CLOSTRIDIUM BOTULINUM  
; TITLE OF INVENTION: NEUROTOXIN  
; NUMBER OF SEQUENCES: 30  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MEDLEN & CARROLL, LLP  
; STREET: 220 MONTGOMERY STREET, SUITE 2200  
; CITY: SAN FRANCISCO  
; STATE: CALIFORNIA  
; COUNTRY: USA  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/405,496A  
; FILING DATE: 16-MAR-1995  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/329,154  
; FILING DATE: 25-OCT-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/161,907  
; FILING DATE: 02-DEC-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/985,321  
; FILING DATE: 04-DEC-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/429,791  
; FILING DATE: 31-OCT-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: INGOLIA, DIANE E.  
; REGISTRATION NUMBER: 40,027  
; REFERENCE/DOCKET NUMBER: OPD-01308  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 705-8410  
; TELEFAX: (415) 397-8338  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2710 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-405-496A-6

Query Match 1.7%; Score 156.5; DB 2; Length 2710;  
Best Local Similarity 18.6%; Pred. No. 0.00065;  
Matches 189; Conservative 154; Mismatches 354; Indels 44;

## RESULT 5

US-08-915-136-6  
; Sequence 6, Application US/08915136  
; Patent No. 6290960

QY 288 KEDIKSHSENVSOHPLHVEV---LHSEIMAHQKFAIRLGSWMNKMINSYSSDPROIFCOAC 344  
Db 355 KSEIFSKLENLNVSDLEIKIAFALGSKI---NOALISKQSGSYLTNLVIEQVKRNYOFLNOH 412  
QY 345 LREEDPSNCLISRLMLWDAKLYGARKILHELIFSSPFMEWEYKKLFAMEFVKYKOL 404  
Db 413 LNPAIESDNN-----FTDTTKIFHDSLFNSATAE---NSMELTKIAPY---L 453  
QY 405 QKEYISDDHDSISITA-----LSVQMFTV-PTL-ARHLIE---EQNVISVITE 448  
Db 454 QVGEMPEARS-TISLSGPGAVASAYDFINLQENTIEKTKASDLIEFKEFPENNLSOLTE 512  
QY 449 TLLEVLPEYLDNRNKNFNQGSQDKLGRVYA-----VICDKLYILISK-PIWITE 497  
Db 513 QEINSLWFDQASAKYQFEKYVRYDRTGGSLSDNGVDENKNTALDKNYLLNNKIPSNVYE 572  
QY 498 RLRMQFLEGFERSFLKILTCMOGMEIRQVQGHIEVDPDWEAAIAIOMOLKNILLMFQEW 557  
Db 573 E-----AGSKNYVHVIIQLQDDDISYEATCNLFKNP--KNSIIQRNMNESAKSY--F 622  
QY 558 CACDEELLVAYKECHKAVMRGCTSFISSTKTVVQSCGH-----SLETKSYR 604  
Db 623 LSDDGESIL---ELNK--YRIPERLKNKEKVKVTFIGHGKDEFTSEFARLSVDSLSNE 676  
QY 605 VSEDLVSIHLPLSRITLHVLRLSRLGAVSRUHEFVSPEDFOVEVLVEPLRCLVLVAQV 664  
Db 677 ISSFLDTIKLIDISPK---NVEVNLGC-----NMFSY-DFNVE--ETYPGKLLLSIMDK 724  
QY 665 VA-----EMWRNGLSLISOVFYQDVQKREEMV--- 693  
Db 725 ITSTLPDVNKNISITIGANOYEVRINSEGRKELLAHSGKWINKEEAIMSDLSKEYIFFDS 784  
QY 694 -----DKDIIMLQIGASLMDPNKLLLVQ-----RYELAE 724  
Db 785 IDNKLKAKSNIPGLASISEDIKTLILDASVDPDKFILNNLKLNESSIGDYIYEKLE 844  
QY 725 AFNKTISTQDQDLIKOYNLTIEMLQVLIYIVGERYVPGVGNVTEETVMTREIHLCLIE 784  
Db 845 PVKNIHNSIDDLIDEFNLL-----ENVSDLEYELKKL----- 877  
QY 785 PMPHSAIAKNLPE-----NENNETGLENNVINKVATFKPGVSGHGVYELKDBSLKD 835  
Db 878 -----NNLDEKYLISFEDISKNNSTYSVRFINK-----SNGESVYVETEKEI--- 919  
QY 836 FNMVEYHYSK-----TQHSKAEHMOKKRRKQENKDEALPPPPPPFCPAFSKV 883  
Db 920 PSKYSEHITKEISTIKNSIITDVNGNLLDNIQDHTSQ----- 957  
QY 884 INLLNCDIMMYILRTVFERAIDTD--SNLWTEGMLQMAFHIALGLLEEKQQLQKAPEE 941  
Db 958 VNTLNA---AFFIQSLIDYSSNKKVDNLDSVSKVQVLAQLFSTGLNTIYDSIQLV--- 1010  
QY 942 VTDFYHKASRLGSSAMN--IOMLEKLEKIGIPQLEGOKDMITW---ILOMFDTVKRLREK 996  
Db 1011 -----NLISNAVNDTINVLPITTEGIPVSTILDGNGLGAARKELLDHDPDLKK 1060  
QY 997 -----SCLIVATTSGSESINKNDEITHDKEAERKKAERKAEARLHROKIMQMS 1043  
Db 1061 ELEAKVGVLAINNSLSIAATVASIVGI-GAEVT-----IFLLPIAGISAGIP 1106  
QY 1044 ALQKNFIETHKLMYDNTSEMPGKEDSIMEEESTPAVSDYSRIALGPKRGPSVTEKE-VLT 1102  
Db 1107 SLVNN-----ELILHDKATSVVNYFNHLSSESKYGLPKTDDDKILV 1147  
QY 1103 CILCOEQEVKIENNANVLSCVQKSTALTQHRKPIELSGEALDPLFMDPDLA 1156  
Db 1148 PIDDLVISEIDFNNSIKLGTG-----NILAMEGGSGHTVTGN-IDHFFSSPSIS 1196



GENERAL INFORMATION:  
APPLICANT: KINK, JOHN A.  
APPLICANT: THALLEY, BRUCE S.  
APPLICANT: PADHYE, NISHA V.  
APPLICANT: FIRCA, JOSEPH R.  
APPLICANT: STAFFORD, DOUGLAS C.  
TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND PREVENTION OF C. DIFFICILE DISEASE  
TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE  
NUMBER OF SEQUENCES: 32  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MEDLEN & CARROLL, LLP  
STREET: 220 MONTGOMERY STREET, SUITE 2200  
CITY: SAN FRANCISCO  
STATE: CALIFORNIA  
COUNTRY: UNITED STATES OF AMERICA  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/915,136  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/480,604  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/405,496  
FILING DATE: 16-MAR-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/329,154  
FILING DATE: 25-OCT-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/161,907  
FILING DATE: 02-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/985,321  
FILING DATE: 04-DEC-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/429,791  
FILING DATE: 31-OCT-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: INGOLIA, DIANE E.  
REGISTRATION NUMBER: 40,027  
REFERENCE/DOCKET NUMBER: OPHD-01763  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 705-8410  
TELEFAX: (415) 397-8338  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2710 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-915-136-6

Query Match 1.7%; Score 156.5; DB 3; Length 2710;  
Best Local Similarity 18.6%; Pred. No. 0.00065;  
Matches 189; Conservative 154; Mismatches 354; Indels 317; Gaps 44;  
QY 288 KEDIKSHSNVQHPHLEVEV---LHSEIMAHQKFAIRLGLGSMWNKIMSYSDSFQIFQAC 344  
DB 355 KSEIFSKLENLVNLSLEKIAFALGVI--NOALISKQGSYTLNVLVIEQVKRYQFLNOH 412  
QY 345 LREEDPSNCLISRLMLDAKLYGARKILHELIFSSPFMEYKKLFAMEFVKYKQL 404  
DB 413 LNPAIESDNN-----FDTTKIFHDSLENSATAE--NSMFLTKIAPY---L 453  
QY 405 QKEYISDDHRSISTA-----LSVQMFTV-PPL-ARHLIE---EQNVISVITE 448  
DB 454 QVGFMPEARS-TISLGGGAVASAYDFINLQENTIEKTLKASDLIEFFKPPENNLSQLTE 512

QY 449 TLLEVLPEYLDNRNKNFQYSDQKLGVRVA-----VICDLKYLISK-PTIWE 497  
DB 513 QEINSLWSDQASAKYQFEKYVRDYTGSGLSSENGVDNFKNTALDKNYLLNKNKIPSNVE 572  
QY 498 RLRMQFLEGFRSFLKILTQMGMEIRROVQGHLEVPDWEAATAIOMOLKNILLMFQEW 557  
DB 573 E-----AGSKNVVHYIIQOGDDISYEATCNLESKNP--KNSIIQRNMNESAKSY--F 622  
QY 558 CACDEILLVAYKECHKAVMRCSTSFISSTKVTVQSCGH-----SLETSYR 604  
DB 623 LSDDGESIL-----ELNK--YRIPERLNKKEKVKVTFICHGKDEFTSEFARLSVDSLSNE 676  
QY 605 VSEDLVSIHPLSRTLGLVRLSRICAVSRLEHFEVSEFEDQVEVLVEPLRCLVLVAQV 664  
DB 677 ISSFLDTIKLIDISPK-----NVEVNLGCG-----NMFYSY-DFNVE--ETYPCKLLSINDK 724  
QY 665 VA-----EMRRNGLSLISQVYFYQDVQKCREMY--- 693  
DB 725 ITSTLPDVNKNSTIGANOYEVIRINSEGRKELLASHGKWINKEAINDLSKSEYIFPDS 784  
QY 694 -----DKDIIMLQIGASLMDPNKFLLLVLQ-----RYELAE 724  
DB 785 IDNKLKAKSNIPGLASISEDIKTLDDASVSPDTKFLNKLNISSIGDIYVEKLE 844  
QY 725 AFNKTISTKDQDLIKQNTLIEEMQLVLIYIGERYVPGVGNVTKEEVTMREIHLICIE 784  
DB 845 PVKNIHNSDDDLIDEFNLL-----ENVSDELYELKLL----- 877  
QY 785 PMPHSAIAKNLPE-----NENNETGLENVINKVATFKKPGVSGHGVYELKDESLKD 835  
DB 878 -----NNLDEKYLISFEDISKNNSTYSVRFINK-----SNGESVYVTEKEI-- 919  
QY 836 FNMIFYHYSK-----TOHSKAEHMKRRKQENKDEALPPPPPPFCFPAFSKV 883  
DB 920 FSKYSEHITKEISTIKNSIITDVNGNLLDNLQDLHTSQ----- 957  
QY 884 INLLNCDDIMYIILRTVFERAIDTD--SNLWTEGMLQMAFHILAGLLEEKQOLKAPEE 941  
DB 958 VNTLNA---AFFIQSLIDYSSNKDVLNDLSVQVLAQVLFSTGLNTIYDSIOLV----- 1010  
QY 942 VTDFYHKASRLGSSAMN--IQMLEKLKIGIPLEGOKDMITW---ILOMEDTVKRLREK 996  
DB 1011 -----NLISNAVNDTINVLPITTEGIPVSTILDGINLGAATKELLDHDPPLKK 1060  
QY 997 -----SCLIVATTSGSEIKNDEITHDKAEKRAKAEARLHRQKIMQMS 1043  
DB 1061 ELEAKVGVLAJNMSLSIAATVASIVGI-GAEVT-----IFLLPIAGISAGIP 1106  
QY 1044 ALQKNFIETHKLMYDNTSEMPGKEDSIMEESTPAVSYSIALGPKRGKPSVTEKE-VLT 1102  
DB 1107 SLVNN-----ELILHDKATSVVNVFNHLSSEKSKYGLPKLTEDDKLV 1147  
QY 1103 CILCQEEQEVKIENNAMVLSACVQKSTALTQHRKPIELSGEALDPLFMDPDLA 1156  
DB 1148 PIDDLVISEIDFNNSIKLGTC-----NILAMEGGSGHTVTGN-IDHFFSPSPSIS 1196

## RESULT 6

US-08-957-310-6  
Sequence 6, Application US/08957310  
Patent No. 6365158  
GENERAL INFORMATION:  
APPLICANT: Williams, James A.  
APPLICANT: Kink, John A.  
TITLE OF INVENTION: IDENTIFICATION OF NEUTRALIZING EPITOPES  
TITLE OF INVENTION: OF TOXIN A AND TOXIN B FOR THE TREATMENT OF C. DIFFICILE  
TITLE OF INVENTION: DISEASE  
NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Medlen & Carroll  
STREET: 220 Montgomery Street, Suite 2200  
CITY: San Francisco

STATE: California  
COUNTRY: United States of America  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/957,310  
FILING DATE: 23-OCT-1997  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/329,154  
FILING DATE: 24-OCT-1994  
APPLICATION NUMBER: US 08/161,907  
FILING DATE: 02-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/985,321  
FILING DATE: 04-DEC-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/429,791  
FILING DATE: 31-OCT-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: Ingolia, Diane E.  
REGISTRATION NUMBER: 40,027  
REFERENCE/DOCKET NUMBER: OPHD-01121  
TELEPHONE: (415) 705-8410  
TELEFAX: (415) 397-8338  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2710 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-957-310-6

Query Match 1.7%; Score 156.5; DB 4; Length 2710;  
Best Local Similarity 18.6%; Pred. No. 0.00065;  
Matches 189; Conservative 154; Mismatches 354; Indels 317; Gaps 44;

Qy 288 KEDIKSHENVSOHPLHVEV---LHSEIMAHQKFAIRLGSMWNKIMSYSSDFRQIFCOAC 344  
Db 355 KSEIFSKLENLVSDLEIKIAFALGVI--NQALISKQSGYLTNLVIEQVKNRQFLNQH 412  
Qy 345 LREPDSNPCLSLRLMDAKLYGARKILHELIFPSFFMEMEYKILFAMEFYKYLQ 404  
Db 413 LNPAGESDNN-----FTDTTKIFHDSLFNSATAE---NSMFLTKIAPY---L 453  
Qy 405 QKEYISDDHRSISITA-----LSVQMFETV-PTL-ARHLIE---EQNVISVITE 448  
Db 454 QVGFMPEARS--TISLSGPGAYASAYDFINQENTIEKTLKASDLIEFKFPENNLSQUTE 512  
Qy 449 TLEVLPEYLDNRNKNFQGSQDKLGRVYA-----VICDLYILISK-PTIWE 497  
Db 513 QEINSLWSFDQASAKYQEKVRYDTGGSLEDNGVDENKNTALDKNVLLNKNKIPSNVE 572  
Qy 498 RLRMQFLEGFRSFKILTCMQMBEIRQVGHIEVDPDPAATAIQMLKNILLMFQEW 557  
Db 573 E-----AGSKNYVHYIIQLQDDISYEATCNLFESKNP--KNSIIQRNMNESAKSY--F 622  
Qy 558 CACDEELLVAYKECHKAVMCRSTSFTSSSTTVVQSCGH-----SLETKSYR 604  
Db 623 LSDDGESIL-----ELNK--YIPRLKNKKEVKVTFIGHGKDEFTSEFARLSVDLSNE 676  
Qy 605 VSDLVSTHPLSLRTLACGLVRLSLGAVSRHLFEVSPEDFQEVVLVEYPLRCLVLVAQV 664  
Db 677 ISSFLDTIKLDISPK-----NVEVNLGC-----NMFSY-DEFNVE---ETYPKLLLSIMDK 724  
Qy 665 VA-----EMWRNRLSLISQVFFYQDVKCREMY--- 693  
Db 725 ITSTLPDVNKNISITIGANOYEVRINSEGRKELLAHSGKWINKKEAIMSDLSKEYIFPDS 784

Qy 694 -----DKDIIMLIQIGASIMDPNPKFLLVLQ-----RYELAE 724  
Db 785 IDNKLKAKSKNIPGLASISEDIKTLTLLDASVSPDTKFIILNNKLNTIESSIGDYIYYEKLE 844  
Qy 725 AFNKTISTKDDOLIKQYNTLIEEMQLVLIYIGERVVPGVGNVTKREVTMRREIHLCLIE 784  
Db 845 PVKNIHNSIDDLIDEFNLL-----ENVSDLEYELKKL----- 877  
Qy 785 PMPHSAIAKNLPE-----NENNETGLENVINKVATFKKPGVSGHGYELKDESLKD 835  
Db 878 -----NNLDEKYLISFEDISKNNSTYSVRFINK-----SNGESVYVETEKEI-- 919  
Qy 836 FNMVYFYHYSK-----TOHSAKAEHMOKKRRKQENKDEALPPPPPEFCPAFSKV 883  
Db 920 FSKYSEHITKEISTIKNSIITDVNGNLLDNIQLDHTSQ----- 957  
Qy 884 INLLNCDIMMYILRTVFERAIDTD--SNLWTEGMLQMAFHIALGALLEKQOQLOKAPEE 941  
Db 958 VNTLNA--AFFIQLSIDYSSNKNLDLSTSVKQVLYAQLFSTGLNTIYDSIQLV---- 1010  
Qy 942 VTFDFYHKASRLGSSAMN--IQMLLEKLGIPQLEGQKDMITW---ILOMFTDVKRLREK 996  
Db 1011 -----NLISNAVNDTINVLPITIEGIPIVTILDINGLGAIAKELLDEHDPDLKK 1060  
Qy 997 -----SCLIVATTSGSESINKDEITHDKAEKRRKKAERKKAARLHRQKINMAOMS 1043  
Db 1061 ELEAKVGVLAINNLSLSIAATVASIVGI-GAEVT-----IFLLPIAGISAGIP 1106  
Qy 1044 ALQKNFIETHKLMYDNTSEMPGKEDSIMEEESTPAVSYSRIALGPKRGPSTYEKE-VLT 1102  
Db 1107 SLVNN-----ELILHDKATSVVNYFNHLSSEKKGYPGLKTEDDKILV 1147  
Qy 1103 CILCOEEQVKIENNAMVLSACVQKSTALTOHRGKPIELSGEALDPLFMDPDLA 1156  
Db 1148 PIDDLVISEIDFNNSIKLGTG----NILAMEGGSGHTVTGN-IDHFFSSPSIS 1196

RESULT 7  
US-10-011-366-6  
Sequence 6, Application US/10011366  
Patent No. 6573003  
GENERAL INFORMATION:  
APPLICANT: Williams, James A.  
Kink, John A.  
TITLE OF INVENTION: IDENTIFICATION OF NEUTRALIZING EPITOPES  
OF TOXIN A AND TOXIN B FOR THE TREATMENT OF C. DIFFICILE  
DISEASE  
NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Medlen & Carroll  
STREET: 220 Montgomery Street, Suite 2200  
CITY: San Francisco  
STATE: California  
COUNTRY: United States of America  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/011,366  
FILING DATE: 16-No. 6573003-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/957,310  
FILING DATE: 23-OCT-1997  
APPLICATION NUMBER: US 08/329,154  
FILING DATE: 24-OCT-1994  
APPLICATION NUMBER: US 08/161,907  
FILING DATE: 02-DEC-1993  
APPLICATION NUMBER: US 07/985,321

```
; ; FILING DATE: 04-DEC-1992
; ; APPLICATION NUMBER: US 07/429,791
; ; FILING DATE: 31-OCT-1989
; ; ATTORNEY/AGENT INFORMATION:
; ; NAME: Ingolia, Diane E.
; ; REGISTRATION NUMBER: 40,027
; ; REFERENCE/DOCKET NUMBER: OPHD-01121
; ; TELECOMMUNICATION INFORMATION:
; ; TELEPHONE: (415) 705-8410
; ; TELEFAX: (415) 397-8338
; ; INFORMATION FOR SEQ ID NO: 6:
; ; SEQUENCE CHARACTERISTICS:
; ; LENGTH: 2710 amino acids
; ; TYPE: amino acid
; ; TOPOLOGY: linear
; ; MOLECULE TYPE: protein
; ; SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-10-011-366-6

Query Match          1.7%   Score 156.5; DB 4; Length 2710;
Best Local Similarity 18.6%; Pred. No. 0.00065;
Matches 189; Conservative 154; Mismatches 354; Indels 317; Gaps 44;

Qy      288 KEDIKSHENVSOHPLHVEV---LHSEIMAHQFPALRLGSGWMNKMISYSDFRQIFCOAC 344
Db      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Qy      355 KSEIFSLENLNVSDLEIKTAFAGLSVI--NQALISKQGSVLTLNVIEQVKRNYOFLNQH 412
Db      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Qy      345 LREEPDSENCLISRLMDWAKLYKGARKILHELFSFFMWEYKKLFAMEFVYYKQL 404
Db      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Qy      413 LNPAIESDNN-----FTDTTKTFHDSLFLNSATAE---NSMFLTKIAPY---L 453
Db      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Qy      405 QKEYISDDHRSISITA-----LSVMQETV-PTL-ARHLE---EQNVISVITE 448
Db      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Qy      454 QVGEMPEARS-TISLSPGAYASAYDFINLQENTIEKTLKASDLIEFKPENNLSQLTE 512
Db      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Qy      449 TLEVLPEYLDRNNKNFNQGYSDKLGRVYA-----VICDLKYIILSK-PTIWTE 497
Db      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Qy      513 QEINSLWSFDQAQSAKYQFEKYVRDYTGGSLSDEONGVDNFKNLTADLNKYNLPKSNNVE 572
Db      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Qy      498 RLRMQFLEGFRSFLKILTCMOGMEIRPROVQGHLEVPDPNEAATAIQMLKNILLMQEW 557
Db      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Qy      573 E-----AGSKNVHYIIQGGDISYEATCNLFKNP--KNSIIHQNMNESAKSY--F 622
Db      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Qy      558 CACDEELLVAYKECHKAVMRCSFTSISSKTVVQSCGH-----SLETKSYR 604
Db      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Qy      623 LSDDGESIL---ELNK--RYPERLNKEKVVFTHGCKDEENTSEFARLSVDLSLNE 676
Db      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Qy      605 VSEDLSVIHLPLRTLAGLVHRLSRGAQSRLHEFVSFEDQFVELVVEYPLRCVLIVVAQV 664
Db      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Qy      677 ISSFLDPTIKLDISPK-----NVEVNLLGC----NMFYSY-DFNVE--ETYPGKLLLSLMDK 724
Db      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Qy      665 VA-----EMWRNRGLSLISQVFFYQDVKCREMY--- 693
Db      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Qy      725 ITSTLPDVNKNSTITIGANQVEVINSGRKELLAHSKWINKKEAIMSDLSUSKEYIFFDS 784
Db      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Qy      694 -----DKDIIMIQIGASIMDNPKELLVLQ-----RYELAE 724
Db      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Qy      785 IDNKIAKSNIPCLASISEDITLLDLDASVPTKFIANNLKNLIESTIGDYIYYEKLE 844
Db      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Qy      725 AFNKTISTKDQDLIKQNTLIEMLQVLIYVGERYVPGVGNVTKEEVTWTREIHLCIE 784
Db      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Qy      845 PVKNIHNISDDLIDEFNLL-----ENVSDLEYELKKL----- 877
Db      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Qy      785 PMPHSATAKNLPE-----NENNETGIENVINKVATPKFGVSGHGVIYELKDLSKD 835
Db      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Qy      878 -----NNLDKYLISFEDIKSNSTSYSVRFINK-----SNGSYVYVETEKEI-- 919
Db      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Qy      836 FNMFYFYHK-----TQHSAEHQMOKRRKQENKDEALPPPPPFCFAPFSKV 883
Db      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Qy      920 FSKYSEHTKEISTIKNSIITDVNGNLLDNIQDHTSQ----- 957
Db      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Qy      884 INLLNCIDIMYILRTVFERAIDTD--SNLWTEGMLOMAFHIALGLGLEERQQOLQAKPEE 941
Db      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
```

```

Db 958 VNTLNA---AFFIQSLIDYSNKNVDVNDLSTSVKVQVLYAQLFSTGLNTIYDSIQLV--- 101
Qy 942 VTFDFYHKASRLGSSAMN--IOMLEKLKIGIPOLEGOKDMITW---ILOMFDTFVKRLREK 996
Db 1011 -----NLISNAVNDINVLPTTTEIGPIVSTTLDGINCAAKEILLDEHDPLKK 1060
Qy 997 -----SCLIVATTSGSEIKNDIEITHDKEKABRKRAKAAARLHROKIMQA 1043
Db 1061 ELEAKVGVLAINMSLSLAATVASIVGI-GAEVT-----IFLLPIAGISAGIP 1106
Qy 1044 ALQKNFTETHKLYMDNTSEMPGKEDSTMESESTPAVSDYSRIALGPKRGPSVTEKE-VLT 1102
Db 1107 SLVN-----ELTHDKATSVVNYFNHLSSESKYKGLKTEDDKILV 1147
Qy 1103 CILCQBQBEVKIENNAVLSCYQKSTALTOHRGKPIELSGEALDPLFMDPDLA 1156
Db 1148 PIDDLVISEIDFNNSIKLGTG----NILAMEGSGHTVTGN-IDHFFSPSPIS 1196

RESULT 8
US-09-091-501B-10
; Sequence 10, Application US/09091501B
; Patent No. 6518413
; GENERAL INFORMATION:
; APPLICANT: Tinsley, Jonathan M
; APPLICANT: Davies, Kay E
; TITLE OF INVENTION: Utrrophin gene expression
; FILE REFERENCE: 620-42
; CURRENT APPLICATION NUMBER: US/09/091,501B
; CURRENT FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: PCT/GB96/03156
; PRIOR FILING DATE: 1996-12-19
; PRIOR APPLICATION NUMBER: GB 9525962.8
; PRIOR FILING DATE: 1995-12-19
; PRIOR APPLICATION NUMBER: GB 9615797.9
; PRIOR FILING DATE: 1996-07-26
; PRIOR APPLICATION NUMBER: GB 9622174.2
; PRIOR FILING DATE: 1996-10-24
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 3433
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (259) ... (250)
; OTHER INFORMATION: Description of Artificial Sequence: Full length
; OTHER INFORMATION: utrophin construct; Xaa = unknown
US-09-091-501B-10

Query Match 1.7%; Score 156; DB 4; Length 3433;
Best Local Similarity 18.1%; Pred. No. 0.0011;
Matches 298; Conservative 259; Mismatches 559; Indels 534; Gaps 80;

Qy 255 DCLABAQLHTTAIDKEGR--RA-----VKAGAVAAQOE-----AKEDIKSHS 295
Db 276 ECEEEEIHQISAVLAEGQSPRATSTVTEVDMDLDSQIALEEVLTWLLSSEDIFQEQ 335
Qy 296 ENYSQHLPHVEYLHSEIMAHQKPAIRLGSMMNKIMSYSDFRQIFQACILREPDSENPC 355
Db 336 DDISD--DVVEKVEQFATHETFMELTAHOSVGSVLAQGNLMTQGTLSBEEPE--- 389
Qy 356 LISRLML---WDAAKLYKGARK--ILHELIFSFFPMEMEYKKLFAME-----FVKYKQL 404
Db 390 IQEQMTLLNARWEALRVESMERGRSLHDAL-----MELOKKQLQQLSSWLTATEERIQM 444
Qy 405 QKEYISDDHRSISITALSVQMFTVPTLARHLIEQNVISVITETLLEVLPEYLDNRNKF 464
Db 445 ESPPLGDD-----LFSLOKLQEHKLSQNDLEAEQVKNSLTHWVIV-----DENSEG 493
Qy 465 NFOGYSQD---KLGRRVAVIC-----DLKYLILISKPTIWTPLRMOFL----- 504

```



Db 725 QMSALQAKLDEEBHKNKLOQHVD-----KLEHSTQMQLFSSSERIDWTQK 771  
QY 561 DEELL---VAYEKCHKAVMRGTSPISSKTVVQSCGHS-----LETYSRVSED 608  
Db 772 QEBLLSQLNVLEKQLOET--QTKNDFLKSEVHDLRVVLSADKELSSVKLEYSFSTQKE 829  
QY 609 -----LVSTHPLSRTLGLHVLRLSRL-----GAVSRLEHVFSE-----DFQ 646  
Db 830 KEFNKLSERHMVQLQDLMLRLENKLESKACQSDYDNLQIMKFEIDQLSRNLQNF 889  
QY 647 VE-VLVEYPLRCLVLVAQVAVEMRRNGLSL-----ISOVFYQDV 686  
Db 890 KENETLSDNLNMLELLE--AEKERNKLSLQFEEDKENSKEILKVLAVRQEKQETA 947  
QY 687 KCREM-----YDKDIIMQIGASLMDPNKFL--LVLRQYELAEAF---NKT 729  
Db 948 KCEQMAKVKLEESILLATEKVISSLE---KSRDSKKVVADLMNQIQELRSSVCEKTET 1004  
QY 730 ISTKQD---DLIKQYNTLI--EEMQLVLIYIGERYVPGVGNVTKEEV---TMRRIHL- 780  
Db 1005 IDTLKQELKDINCKYNSALVDREESRVLI-----KKQEVLDLDELKTLRLR 1050  
QY 781 -----LCIEPMPHSAIAKN-LPENENNETGL-----ENVINKVATFKPGVSGHG 824  
Db 1051 ILSEDIERDMLC-EDLAHAQEQLNMLTEASKKHSLQSAQBELTKKEALIQE---LQHK 1106  
QY 825 VYELKDESUKDFNMYFYHYSKTOH---SKAEHMOKRRKROENKDEALPPPPPPFCPAPS 881  
Db 1107 LNOKKEVEQKKNYFNKMRQLEHVMSAAEDPOS-----PKTPPHFQTHLA 1153  
QY 882 KVINLLNCDIM-MYILRTVFERAI---DTDSNLWTEGMLQMAFHILALGLLEEQLOKA 937  
Db 1154 KLETOEQEIEDGRASKTSLEHLVTKLNEDREVKNAILRM-----KEQLE- 1200  
QY 938 PEEVTFDFYHKASRLGSSAMNIQMLLEKLGIPQLEGQKDMITWILQ-MFDTVKRLREK 996  
Db 1201 -----MENLRLESQOLIEK-----NWLQGLQDIDIKROKE- 1230  
QY 997 SCLIVATTSSESINKDEITHDKAEKRRKAEARLHROKIMAQMSALQKNFTHTKLM 1056  
Db 1231 -----NSDONHPDNOOLKNEQESIKE-----RLAKSKIVEMLKMKADLEEVQSAL 1277

## RESULT 10

US-09-723-262-2

; Sequence 2, Application US/09723262  
; Patent No. 6379912  
; GENERAL INFORMATION:  
; APPLICANT: Beraud, Christophe  
; APPLICANT: Sakowicz, Roman  
; APPLICANT: Wood, Kenneth  
; TITLE OF INVENTION: No. 6379912el motor proteins and methods for  
; FILE OF INVENTION: their use  
; FILE REFERENCE: 1017  
; CURRENT APPLICATION NUMBER: US/09/723,262  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: US 09/572,191  
; PRIOR FILING DATE: 2000-05-17  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 1388  
; TYPE: PRT  
; ORGANISM: Human  
US-09-723-262-2

Query Match 1.6%, Score 150.5, DB 4; Length 1388;  
Best Local Similarity 19.3%, Pred. No. 0.00073;  
Matches 212; Conservative 162; Mismatches 372; Indels 355; Gaps 54;  
QY 187 LNEEVIQARKIPSPVYKIVVEMTWEE---KE-----LPELQIREKNERCYCVLFND- 238  
Db 447 LKREKFIQSNKM---IVKRFEDQIIRLEKLHESRGGFLP-----EQDRLLSERNEI 497  
QY 239 -----EHSYDHVIYSLQALCELAELHTTAIDKGRRAVKAGAYAACQAKEDI 291  
Db 498 QTLREQLIEHP-----RVAKYAMENHSLREENRL-----RLLEPV 533  
QY 292 KSHSENVSQHPLHVEVLHSEIMAHQKPAIRLGWSMNKIMSYSSDFRQIFQACLRPEPDS 351  
Db 534 KRAQEMDAQTIAKLEKAFSEISGMEK-----SDKNOQGFSPKAQKE----- 574  
QY 352 ENCLLSRLMDAKLYGARKILHELIFSSFFMEMEYKILF---AMEFYVYKQLOKEY 408  
Db 575 --PCLFANTEKLAQALL---QIOTELNNSKQEYE-EFKELTRKROLESELOSLQAN 627  
QY 409 IS-----DDHRSISITALSVMFTVPTLARHLIEEQNVISVITELLEVLP 455  
Db 628 LNLLENLEATKAKRQEVSQLNKLHAETKLIITPTKAYQL-----HSRPVKLSP 678  
QY 456 EYLDNRNKNFQGYSDQKLG RVY---AVICDLKYILISKPTIWTIRLMQFLEGRFSPLK 512  
Db 679 E-----MGSGSLYTONSILLNDILNEPVPPEMNEQAFEAISEELRTVQE 724  
QY 513 ILTCMQM---PEIRROVQOHIEVDPDWEAAIAIQOLKNILLMFQ-----WCAC 560  
Db 725 QMSALQAKLDEEBHKNKLOQHVD-----KLEHSTQMQLFSSSERIDWTQK 771  
QY 561 DEELL---VAYEKCHKAVMRGTSPISSKTVVQSCGHS-----LETYSRVSED 608  
Db 772 QEBLLSQLNVLEKQLOET--QTKNDFLKSEVHDLRVVLSADKELSSVKLEYSFSTQKE 829  
QY 609 -----LVSTHPLSRTLGLHVLRLSRL-----GAVSRLEHVFSE-----DFQ 646  
Db 830 KEFNKLSERHMVQLQDLMLRLENKLESKACQSDYDNLQIMKFEIDQLSRNLQNF 889  
QY 647 VE-VLVEYPLRCLVLVAQVAVEMRRNGLSL-----ISOVFYQDV 686  
Db 890 KENETLSDNLNMLELLE--AEKERNKLSLQFEEDKENSKEILKVLAVRQEKQETA 947  
QY 687 KCREM-----YDKDIIMQIGASLMDPNKFL--LVLRQYELAEAF---NKT 729  
Db 948 KCEQMAKVKLEESILLATEKVISSLE---KSRDSKKVVADLMNQIQELRSSVCEKTET 1004  
QY 730 ISTKQD---DLIKQYNTLI--EEMQLVLIYIGERYVPGVGNVTKEEV---TMRRIHL- 780  
Db 1005 IDTLKQELKDINCKYNSALVDREESRVLI-----KKQEVLDLDELKTLRLR 1050  
QY 781 -----LCIEPMPHSAIAKN-LPENENNETGL-----ENVINKVATFKPGVSGHG 824  
Db 1051 ILSEDIERDMLC-EDLAHAQEQLNMLTEASKKHSLQSAQBELTKKEALIQE---LQHK 1106  
QY 825 VYELKDESUKDFNMYFYHYSKTOH---SKAEHMOKRRKROENKDEALPPPPPPFCPAPS 881  
Db 1107 LNOKKEVEQKKNYFNKMRQLEHVMSAAEDPOS-----PKTPPHFQTHLA 1153  
QY 882 KVINLLNCDIM-MYILRTVFERAI---DTDSNLWTEGMLQMAFHILALGLLEEQLOKA 937  
Db 1154 KLETOEQEIEDGRASKTSLEHLVTKLNEDREVKNAILRM-----KEQLE- 1200  
QY 938 PEEVTFDFYHKASRLGSSAMNIQMLLEKLGIPQLEGQKDMITWILQ-MFDTVKRLREK 996  
Db 1201 -----MENLRLESQOLIEK-----NWLQGLQDIDIKROKE- 1230  
QY 997 SCLIVATTSSESINKDEITHDKAEKRRKAEARLHROKIMAQMSALQKNFTHTKLM 1056  
Db 1231 -----NSDONHPDNOOLKNEQESIKE-----RLAKSKIVEMLKMKADLEEVQSAL 1277



; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3248 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: HUMAN  
; US-08-353-700-1

Query Match 1.6%; Score 147.5; DB 1; Length 3248;  
Best Local Similarity 18.3%; Pred. No. 0.0059;  
Matches 318; Conservative 253; Mismatches 632; Indels 537; Gaps 81;  
QY 171 HEPGRAGTIGKNSRCPNPEE---VIVQARKIFPSPYKIVVEMTWEEB---KELPPELQ 223  
DB 304 HEKEMKGVNKTQELQLEKAKVELIEKEVNLKCRDELVRTTAQYDQASTKYTALEQK 363  
QY 224 TREKNERYCVLFNDE--HHSDYHVIYSLQALDCELAQAHLHTTAIDREGRRAVAGAY 281  
DB 364 LKKLTEDLSQORNAESARCSLEQKIKEKEFEQELSQRQSFQTLDOECIQ-MKARLT 422  
QY 282 AACQAKEDIKSHSENVSOPHLVHEVSEIWAHQFALRLGSMWNKINSYSSDFROIFC 341  
DB 423 QELOQAKN-----MH-NVLQAEID-----KLTSVKQOLENNLEEFKQKLC 461  
QY 342 QACLRPEPDPENPCLLISRLMLDWAQKILHKGARKILHELIFSSPFMEYKKLFAFEVKKY 401  
DB 462 RA-----EQAFQASQIKENELRR--SMEEMKE 487  
QY 402 KOLQKEYISDDHRSISITALSQVMTVP---TLARHLIEBQNVISVITETLEVLPEYL 458  
DB 488 NNLLK---SHSQKAREVCHLAELKNIKQCLNQSQNFABEMKAKTSQETMLRDQEE-- 542  
QY 459 DRNNKFNFGYSQDKLGRVYVYICDLKYILISKPTIWTERRMQFLEGRSFILKILTCMQ 518  
DB 543 ----KINQENSL-TLEKLKLAADLE-----KQRCDSQ 571  
QY 519 GMEETRRQVGHIE-----VDPDWEAAI-AIQM-----QLKNILLMFQWCADEE 563  
DB 572 DULKKRE---HHIEQNDKLSKTEKESKALLSALELKKEYBELKEETLFCQWKSNEK 628  
QY 564 LLLVAYKECHKAVMRCSTSFSSKTVQSCGHSLETKSYRVEDLSVSIHLP---LSRTL 620  
DB 629 LL-----TQMESEKENLOSKINHLETCCLKTQIKSHYENRVRTLEMDRENLSVEI 679  
QY 621 AGLH-----VRLSRLGAVSRHLHEFVSFEDQVEVLVE-YPLRCLVLVLAQVVAEMWR 670  
DB 680 RNLHNVLDKSVSEVETQKL-AYMELQQAEEFSQKHQKEIENMCLNKLTSQITQGVED--L 735  
QY 671 RNLGLSISOVFFYQDKCREMYD-----KDIIMLQIGASLM---DPNKLFLVLRQYEL 722  
DB 736 EHKLOLLSNEIMDKD-RCYQDLHAEEVSLRLDLKSK-DASLVNEDHQRLSLAFAQOQAM 793  
QY 723 AEAFNKTI-----STKQDOLI-----KOYNTLIEEML 749  
DB 794 HHSFANIIGEQQSMPSERSECRLEADQSPKNSAILQNRVDSLEFSLQSKQMNSDLOKQC 853  
QY 750 QVLIYIVGGRYPVGVNVTKEEVTMRHETIHLICI-----EPMHP-SAIANKLP 796  
DB 854 EELVQIKGE-----IEENLMKAEQHQSVFVASTSQRISKLOEDTSAHQNVVAETLS 904  
QY 797 ENENNETGLYENKIVAT-----FKKPGVSGHYELKDESLKDFNMFFYHYSKTQHS 849  
DB 905 ALENKEKEQLQNDKVEVTEQAELQELK---SNH-----LLEDLSKELQLL-----S 948  
QY 850 KAEHQKRRKQENKDEALPPPPPPFCFAPAFSKVINLNCIDIMMYILRTVTFERAID--TD 907  
DB 949 ETLSLSEKKE-----MSSIISL-----NKREIBELTQ 974

QY 908 SNLWTEGMLQWAFHILALGILLEEKQOQLOKAPBEEVTFDFYHKASRLGSSAMNIQMLLEKL 967  
DB 975 EN-----GTLKE-----INASLQERKMNIQKSE---SFANYIDEREKSISELSQYKQEKL 1023  
QY 968 KGIPLEGQKDMITWILQMFDTVKRLREK-SCLIVATTSGSESIKND---EITHDKEKA 1022  
DB 1024 ILQRCETGNAYEDLSQYKAAQEKNSKLECLNNECTSLCENRKNKELEQLKEAKAHEQ 1083  
QY 1023 ERKKAABAARLHRQKIMAOVSALQKNFTETHKLMYDNTSEMPGKEDSIMEESTPAVS DY 1082  
DB 1084 EFLTKLAFABERNQNLMELETVOQ---ALRSEMTDNQNSKSEAGGLKQEIEMT----- 1134  
QY 1083 SRTALGCRKGRPSYTEKEVLTICLQEBQEVKIENNAMVLSACVQKSTALTQHRGKPIELS 1142  
DB 1135 -----LKEQKNQKQEV-NDDLQENELQMKVMK-----TKHE----- 1165  
QY 1143 GEALDPLFMDPLAYGTYTGSCGHVMAVQKFEAVOLSSQQRHIVDLDFLEGEVLC 1202  
DB 1166 -----CONLESEPIRNSVKER-----ESE----- 1184  
QY 1203 PLCKSLCNTVPIPIIPLQPKINSNADA-LAQLLTLLAR-----WIQT 1243  
DB 1185 ---RNQCN-FKPQMDLEVKEISLDSYNAQLVQLEAMLRNKLKQSEKECECLQHELQT 1240  
QY 1244 VLARISYNTIRHAKGNPIPIFPNOGMDSTLEFHSILSFGVSESIKISYNIKEMVILFA 1303  
DB 1241 IRGDLTSLNLDQMSQE-----ISGLKDCIDAEEKYISGPHELSTSQNDNAHLQCSIQ 1294  
QY 1304 TTIYRIGLKVPPDERDPRVPMLTWSTCAFTIQTATENLGLDGEKPLFGALQNRHQHGLAL 1363  
DB 1295 TTMNKLN-----ELEKICEILOAEKYEVLTELNDSRSCITAT 1332  
QY 1364 MQFAVAGRITCPQVLIQHLVRLSVLPNTKSEDTPCLLSIDLPHVLVGAFLAPSLYW 1423  
DB 1333 RKMA-----EEVGKLLNEV--KILNDDSG-LLHGEVLEDIPGGEFG-EQPN 1375  
QY 1424 DDPVDLPSSVSSYNHLYLPHLITMAHMLQIILTVDTGLPLAQVOEDESEAH-SASFF 1482  
DB 1376 QHPVSLAPLDESNEYHLTISDKQVQMFAEL---QEFSLSLQSEHKILHQHCQMSKK 1432  
QY 1483 AEISOYTSG-----SIGCDIPGWY-----LWVSLKNGITPYLRCAAL----- 1519  
DB 1433 SELQTVVDSLKAENLVSTNLNRNFQGLVKEQMLEGLVPSLSSSCVPDSSSSUSSLD 1492  
QY 1520 --FFHYLLGVTPPEELHTNSAEGEYSA-----LCSYLS-----LPTNLF 1557  
DB 1493 SSFYRALLEQTGDMSLLSN-LEGAVSANQCSVDVFCSSLQTYVYVDSLKAENLVSTNL-R 1550  
QY 1558 LFOEYWDTVR-----PLIQRCA-DPALLNCLKQKNTVVYVYPRKRNLSIELPDD 1605  
DB 1551 NFQ-GDLVKEMQLGLEGLVPSLSSSCVPDSSSSLSLGDSSFY-----RALLEQTGD 1601  
QY 1606 YSCLLNQASHFRCPRSADDERKHVPVLCFLCAILCSQNICQCOEIVNGEVEGACIFALHLC 1665  
DB 1602 MSLSNLNLEGVYVANSQCSDVE-----VFCSS-LQEENLTRKETPSAPAKGVVEELES-- 1651  
QY 1666 KARGCAYPAPVDEYGE--TDPGLKRGNPLH-----LSRER-----YRKLHL-----VMOQ 1709  
DB 1652 ----CEVYROSLEKEEKEMESQGIKMKNEIOELQLLSERQELDCLRKQYLSENEQWOO 1707

## RESULT 13

PCT-US95-16216-1  
; Sequence 1, Application PC/TUS9516216  
; GENERAL INFORMATION:  
; APPLICANT: Yen, Timothy J.  
; APPLICANT: Rattner, Jerome B.  
; TITLE OF INVENTION: Nucleic Acid Encoding a Transiently  
; TITLE OF INVENTION: Expressed Kinetochores Protein, and Methods of Use  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dann, Dorfman, Herrell and Skillman  
; STREET: 1601 Market Street Suite 720





Db 1602 MLLSLNLEGVWSANQCSVDE-----VFCSS-LOENLTKRETSPAPAKGVLEESL-- 1651  
QY 1666 KARGCAYPAPYDEYGE--TDGLKRGNDPLH-----LSRER-----YRKLHL-----VWOO 1709  
Db 1652 ----CEVYRQSLEKLEKMEGOGIMKNKEIQEQLQLSSERQELDLCKRQYLSENEQWOO 1707  
RESULT 14  
US-09-914-259-67  
; Sequence 67, Application US/09914259  
; Patent No. 6495336  
; GENERAL INFORMATION:  
; APPLICANT: Makowski, Lee  
; APPLICANT: Hyman, Paul  
; APPLICANT: Williams, Mark  
; TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES  
; FILE REFERENCE: 8471-010-999  
; CURRENT APPLICATION NUMBER: US/09/914,259  
; CURRENT FILING DATE: 2000-11-21  
; NUMBER OF SEQ ID NOS: 180  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 67  
; LENGTH: 959  
; TYPE: PRT  
; ORGANISM: Rattus norvegicus  
; US-09-914-259-67

Query Match 1.5%; Score 140.5; DB 4; Length 959;  
Best Local Similarity 20.6%; Pred. No. 0.0031;  
Matches 155; Conservative 112; Mismatches 256; Indels 231; Gaps 37;  
QY 526 QVGQHIEVDPPWEAIAIQMOLKNILLMFQEWACADELLLY-----AYKECHKAV 576  
Db 264 EVG---DENPCWSA-----QKVTNLHLMQ-----LVRVLVSPTPPPGATSSCQKAM 307  
QY 577 MRCS-----TSFSSKTVVQSGHSLSETKRYRVED-LVSIHLPL 616  
Db 308 FQCGLLQQLCTILMATGIPADILTTINTVSEVIRG-----QVNQDYFASVNA 358  
QY 617 SRTLAGLHVRSLRGAVSRLEHFFSVFQEVVLPCLVLAQVVAEMWRNGLSL 676  
Db 359 NPPRAIVVLL--MSWNERPPFV-----LRCAVLYC----- 388  
QY 677 ISOVFYODVCKREMYDKDIIMLQIGAS-----LMDPNKELLLVLQRY-----ELAE 724  
Db 389 -FQCLYKNEQGQEV-ATLLPSTIDATGNSVSAGQLCGGLFSTDSLSNCAVALAH 446  
QY 725 AFNKTISTKQDLIKQYNTLI-----EEMQLVLIYV--GERVYVPGVN-----VTKEE 771  
Db 447 ALOGNATQKQLLRVOLATSGNPPVSLLOQCTNLSQGSKIQTRVGLLMLLCTWLSNCP 506  
QY 772 VTMREIHLICIEPMPHSAIAKNLPENNETGLENVINKVATFKKPGVSGHGVYELKDE 831  
Db 507 IAVTHFLHNSANVPPLTQGAENLGEELQVQGLCALL-----LGTSIY-FNDN 554  
QY 832 SLKDFNMFYHYSKTQHSKAEHMQKRRKOENKDEAL-----PPP-PPP 874  
Db 555 SLE-----NYTK-----EKLQOLIEKRIGKENYIEKLGFIKSHLYSRASQRPQPNFPSP 604  
QY 875 E---FCPAFSKVINLNCIDIMYILRTVFE-----RAITDSNLTWTEGLQWAFHIL 923  
Db 605 EYMFDPHFETKLVELEGVITKAIYKSSSEEDKKEEVKKTLEOHNVITHYKNMTRQDL 564  
QY 924 ALGLEEK-----QOLQKAPEEVT-----FDFYHKASRLGSSAM 958  
Db 665 QLEELKQVSTLKCQNEQLQTAVTQOASIQQHKDQYNLLKVQLGKDNHGHQSHSDGAQV 724  
QY 959 N-----TOMLEKLUKGPQEGQKDMTWIQLNFDTVKRLREKSLCIIVATTSGSEIKN 1012  
Db 725 NGIOPEEISRLREE---IEELRSHQVLLQ-----SOLAEEKDTVI-----ENLRS 765  
QY 1013 DEITHDKAEK---RKRKAEARLHRQKIMQMSALQKNFIETHKLMYDNTSEMPGKEDS 1069

Db 766 SQVSGMSEQALATWCSPRDAEQVABLQKELSALKSOLCSQSLEITRLQTEN-SELOQRAET 824  
QY 1070 IMBEEESTPAYSDYSRIALGPKRGPVTEKEVLTILCQEEQEVKIENNANWVLSACVQKST 1129  
Db 825 L--AKSVP-VEGESLVTAAK---TTDVEGRLSALLQETKELKNE-----IKALSEERT 872  
QY 1130 ALTOHRGCKPIELSGEALDPLFMDPDLAYGTYTGS 1163  
Db 873 AIQ-----KQLDSSNSTIAILOTEKDKLYLEVTD 902  
RESULT 15  
US-08-493-092-2  
; Sequence 2, Application US/08493092  
; Patent No. 5728807  
; GENERAL INFORMATION:  
; APPLICANT: Shiloh, Yosef  
; APPLICANT: Tagle, Danilo A.  
; APPLICANT: Collins, Francis S.  
; TITLE OF INVENTION: Ataxia-Telangiectasia Gene  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Reising, Ethington, Barnard & Perry  
; STREET: P.O. Box 4390  
; CITY: Troy  
; STATE: Michigan  
; COUNTRY: US  
; ZIP: 48099  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/493,092  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Kohn, Kenneth I.  
; REGISTRATION NUMBER: 30,955  
; REFERENCE/DOCKET NUMBER: P-310 (TAV)  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (810) 689-3500  
; TELEFAX: (810) 689-4071  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1708 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-08-493-092-2

Query Match 1.5%; Score 140.5; DB 1; Length 1708;  
Best Local Similarity 17.9%; Pred. No. 0.0083;  
Matches 260; Conservative 197; Mismatches 469; Indels 529; Gaps 68;  
QY 324 SWNKMYSYSSDFRQIFCQACLRB-----EPDSENP-----CL-----ISRLMLWD 364  
Db 509 SWRNLLSTHV---QGFFTSCLRHFSQTSRSTTTPANLDSSEHFFRCCLDKKSQRTMLAV 564  
QY 365 AKLYKGARKLHELIES-SFFEMEXKKL-----FAMEFVYKYLQKEYISDD 412  
Db 565 VDMRQKRQSSQSTTNDFAFWLDLNYLEVAKVAQSCAAHFTALLYAEIY---ADKKSMDDQ 622  
QY 413 HDRSISITALSVMQFVPTLRLHIEQNVISVITETLLLEVLPEYLDNRNNKFNFGYSQD 472  
Db 623 EKSLAFEEGS-QSTISSISEKSEETGI---SLQDILLEIYRIGPDSLYGCGG----- 675  
QY 473 KLG RVYAVICDLKYILLISKPTIWRRLRMQFLEGRSFLKTLTCMQGMEETRRVQGO-HI 531  
Db 676 --CKMLQPTITRLR--TYEHEAMWKGALVTVDLE---TAIPSSRQAGTIOALQNLGLCHI 728

Qy	532	EVPDPWEAAIAIQMLKNIILMFQEWACDELLVA-----YKECHKAVMRC-STSF	584
Db	729	-----LSVYLGGLDYENKDMCPLEELHLYQAARWQMWDHCTSVSKEVEGYSIH	777
Qy	585	SSSKTVVQSCG-----HSLETKSVYSEDILVSIHLPLSRLAGLHVLRSLRGAVSRL	636
Db	778	ESLYNALQSURDREFSTFYESLKYARKEYEEMCKRSL-----ESVSYLPTLSLRQAIGEL	834
Qy	637	HEFVSFEDFOV-----EVLVEYPLRCLVLVAQVVAEMWRNRGLSLISQVFFQDV---	686
Db	835	ESTIGELFSRVTHROUSEVIK-----WOKHSQLLKDSFSEQPTMAL	878
Qy	687	-----KCREEMYDKDIIMQIGASLMDPNKFLLLVLQRYELAEAFNK	728
Db	879	RTVILEITMEKEMDNSORECIKDILTKHLVELSI-----LARTFKN	919
Qy	729	T-ISTRDODLIIKOYNL-----TEEMLOVLIYIVGERYPGVGNVTYKEEVTMRIIH	779
Db	920	TQUPERAIFQIKOYNSVSCGVSEWQLEE-AQVF-----WAKKEOSLASLIL	964
Qy	780	LLCIEPMPHSAIAKN-----LPENENNETHLENVINKVATFFKPGVSGHGV-	826
Db	965	KOMIKKLDASCANNPSLKITYTECLRYCGNWLAETCLENPVIMQTYLEKAVEVAGNYD	1024
Qy	827	-----ELKDSLKDFNMFYHYSKTQHSKAEHMQK-----KRRKOE-----	862
Db	1025	GESSDELNRNGMKMAF-LSLARFSDTOYORTENYMKSEFENKQALLKRAKEVEGLLREHK	1083
Qy	863	-----NKDE-----ALPPPPPEFCAPSKVINLNCDIM-----MYILRTV	899
Db	1084	IQTNRVTVKVORELEDELALRALKEDKRFLCKA---VENYINCLLSGEEHDMWVFL-	1139
Qy	900	FERAIDTDSNLWTEGMLQMAFIALGLLEEKQOLOKAPEEYTFDP-----YHKASPLGS	955
Db	1140	-----CSLWLENS-----GVSVNGMKMRDGMKIPTYKPLMYQLAARNGT	1181
Qy	956	SAMNIQMLEKLGIPQLEGOKDMITWLOMFDTVKRLREKSLDIVATTSGSESIKNDEI	1015
Db	1182	KMWG-----GLGFHEVLNLSRI-----SMDPHHTLFIILALANAN---RDEF	1223
Qy	1016	THDKEKAERK-----KAEAA-----RLHROKIMQMSALOKNFI---	1050
Db	1224	LTRPEVARRSRITKNNVPKQSSQLEDDETEAANRIICTIRSRRPQMRSVREALDAYILA	1283
Qy	1051	-----ETHKLMYDNTSEMPGKEDSTIMEESTPAV-----SDYS-----R	1084
Db	1284	NLDATOWKTORKGINPADQPIYTKLNLIEDVVVPTWEIKVDHTGEVGNLVTIOSFKAEFR	1343
Qy	1085	IAGPKRGPVSVEKVLTCI-----KAEAA-----LCOEEQEVKIEIN	1116
Db	1344	LAGG-----VNLPKIIDCVGSDGKERRQLVGRDDLQDQAVMQVFMCMNTLLQRNTET	1397
Qy	1117	NAMVLSACVOKSTALTQHRG-----KPIELSGEALDPLFMDPDLAYTITYGSCGHVM	1168
Db	1398	RKRKLITCTYKVVPLSORSQVLEWCTGTVPi---GEF---LVANEDGAHKRYRPNDPSAF	1451
Qy	1169	HAYCWOXYFEAVOLSSOORTHVDLFDESLEYCLPCKSLCNTVITPIIPLQPQKINSENA	1228
Db	1452	Q---CQKMWVEQKSEKYEVEFM-----DVCQNFQPVF-----RY	1485
Qy	1229	DALAQLTLTARWIQTVLA-----RISGYNIRHAKGENPIPIFFMQMGDSTLEPHSI	1280
Db	1486	FCMEKFLDPAIWEKRLAYTRSVATSIYGYL-----GLGDRHVQ-----	1526
Qy	1281	LSFGVRESSIKYNSIKEMWILFATTYIRIGLYVPPDERP-----RVPMLT	1326
Db	1527	-----NILINQSAELHIDLGVAPEOGKIILPTETVFRLTRDITVDGMGITGVGEVF	1579
Qy	1327	WSTCAFTIOAIEN-----LLGDECKPLFGALONQHGLKALMQFAVAQRITCP	1375
Db	1580	PROCKETMYWRNSQETLLTITVEVLLYD---PLF---DWTMPLKALY---LOQR---P	1626
Qy	1376	QVLIQHLVRLLSVLPFNKISDPTCLLSIDLFLHVLVGAFLAFPSLWDDVDVLQPSVSV	1435

Search completed: September 25, 2003, 14:46:02  
Job time : 37 secs

1	2612.5	28.6	1109	10	US-09-529-063-55	Sequence 55, Appl
2	2612.5	28.6	1109	12	US-10-414-378-55	Sequence 55, Appl
3	250	2.7	247	15	US-10-050-704-110	Sequence 110, Appl
4	184	2.0	1979	15	US-10-205-823-419	Sequence 419, Appl
5	180	2.0	996	9	US-09-815-242-5251	Sequence 5251, Ap
6	180	2.0	1009	9	US-09-815-242-12141	Sequence 12141, A
7	175	1.9	2383	15	US-10-082-830-260	Sequence 260, App
8	168.5	1.8	3899	15	US-10-171-311-4	Sequence 4, Appli
9	168.5	1.8	3907	15	US-10-171-311-2	Sequence 2, Appli
10	168.5	1.8	3917	15	US-10-171-311-8	Sequence 8, Appli
11	168.5	1.8	3925	15	US-10-171-311-6	Sequence 6, Appli
12	165	1.8	2649	12	US-10-205-219-169	Sequence 169, App
13	164.5	1.8	2139	9	US-09-727-384-6	Sequence 6, Appli
14	164.5	1.8	2139	15	US-10-023-219-4	Sequence 4, Appli
15	156.5	1.7	2710	15	US-10-011-366-6	Sequence 6, Appli

Db 121 PGVQVNADEIRKRIIHLQISKPMHSELVKSIPEDENKETGMESVIEAVAHFKKGLT 180  
Qy 822 GHGVYELKDSKDFNMVYHYSKTQHSKAEHMQRKRKQENKDEALPPPPPPFCFPAFS 881  
Db 181 GRGMVELKPECAKEFNLYFHFSAEQSKAEAAQRKLRQNRDPTALPPVLPFCPLFA 240  
Qy 882 KVINLLNCDIMMYILRTVFERAIDTDSNLWTEGMLQMAFHILALGLLEEKQOLKABEE 941  
Db 241 SLVNILQSDVLMCTMGITLOWAVEHNGYAWSESMQLQRLVHLHIGMALQEEKHLENVTEH 300  
Qy 942 -VTFDFYHKASRLGSSAMN---IQMLEKLGIPQLESGQDMITWILQFDTVKRLREKS 997  
Db 301 VVTFFTQKISKPGAPKNSPSILAMLETQNAPLYEVHDMIRWILKTNVAVKMKRESS 360  
Qy 998 -CLIVATTSGSEIKNDEITHDKAEKRRKAEARLHROKIMAOQSALQKNFTETKLM 1056  
Db 361 PTPSPAETEGTIM---EESRDKDAERKAEARLRRKIMAOSEMQRHFIDENKEL 417  
Qy 1057 YDNTEMPGKEDSIMEEESTPAVSYSRIALGKRGPSVTEKEVLTICILQOEQEVKIE 1116  
Db 418 FOOTLELDASTSAVLDH--SPVASDMTLTALGPAQTVPEQRFVTCILQOEQEVKIES 475  
Qy 1117 NAMVLSACVOKSTALTQHRKPIELSGEALDPLFMDPLAYGTGSCGHVMAHVCWQY 1176  
Db 476 RAMVLAFAVQRTSVLSKNRSKFQ-DPEKYDPLFMHDPDLSCGTHSCGHIMHAHCWQY 534  
Qy 1177 FEAVOLSSQO-----RIHVDLDESGEYLCPLKSLCNTVPIPILOPOKINSADAL 1231  
Db 535 FDSVQAEQRQRLRLHTS-YDVENGFEPLCECLSNVPIPL-LSPRINFNNRLN-F 591  
Qy 1232 AQLTLARWOTVLARISGYNIRHAKENPIPIFFNQGMDSTLEFHSILSFGVSESIKY 1291  
Db 592 SDQPNLTQWIRTISQIKALQFLRKEESTPNNASTKNSENVDLQLPPEGFRDPKIPY 651  
Qy 1292 SNSIKEMVILFATTYRIGLVKPPDRDPVRPMLTWSTCAFTQAIENLLGDEKGPLFGA 1351  
Db 652 SESIKEMLTFTGTATYKVLGVHPNEEDPRVIMCWGSCAYTIQSIERILSDEKPLFGP 711  
Qy 1352 LQNRQHGLKALMQFAVQRTCPQVLOIQRHLLSVLPNISKEDTPCLLSLDLHVL 1411  
Db 712 LPCRDLCLSLTRFAAAHWTVASVVOGHFCLFASLPVNDSSHEELPCILDIDMHLL 771  
Qy 1412 VGAVLAPPSLYWDVDPDLQPSVSSYNNHLYLFHLITMAHMLQILL---TVDTGILPAQV 1468  
Db 772 VGLVLAFLAQCD----PSGISLGTGLHIFHLVTHMAHIQILLTSCTEENGMD--QE 824  
Qy 1469 OEDSEEAHSASSFAEISQVTSIGCDIP-GWYLVWSLKNGITPYLRCAALFPHYLLGV 1527  
Db 825 NPPOEESAVLALYKTLHQYT-GSALKEIPSGHNLWSVRAGIMPFLKCSALFFHYLNGV 883  
Qy 1528 TPPEELHTNSAGEYSALCSVLSLPTNLLFQBYWDTVTRPQLLQRRCADPALLNCKOKN 1587  
Db 884 PSPDIOV-PGTSFHEQLCSYLSLPNNLICLFQENSEIMNSIESWCNSEVKKYLEGER 942  
Qy 1588 TVVRYPKRNSLIELPDYSCLLNQASHFRCPRSADDERKHPVLCFLCGAILCSONICCO 1647  
Db 943 DAIRYPRESKNLNPEDYSLINQASNFSPCKSGGDKSRAPTLCLVCGSLCSQSYCCQ 1002  
Qy 1648 EIVNGEEVGACIFHALHC-----KARGCAYPAPYLDYEGETDPLG 1687  
Db 1003 TELGEDVGACHTATYSCGSGVGIFLVRRECQVFLAGKTKGCFSPYLDYGETDQGL 1062  
Qy 1688 KRGNPLHLRSRYRKLHVWQOHCIIIEIARQSQTNQMLFGFNWOLL 1734  
Db 1063 RRGNPLHLCKERFKKIQKLHQHVSVEIGHAQEAQNTLVGIDWQHL 1109

## RESULT 2

US-10-414-378-55

; Sequence 55, Application US/10414378

; Publication No. US20030165981A1

; GENERAL INFORMATION:

;

; APPLICANT: FUKUSHIMA, DAIKICHI  
; APPLICANT: SHIBAYAMA, SHIRO  
; APPLICANT: TADA, HIDEAKI  
; TITLE OF INVENTION: POLYPEPTIDE, CDNA ENCODING THE POLYPEPTIDE, AND USE OF  
; TITLE OF INVENTION: THE BOTH  
; FILE REFERENCE: Q58769  
; CURRENT APPLICATION NUMBER: US/10/414,378  
; CURRENT FILING DATE: 2003-04-16  
; PRIOR APPLICATION NUMBER: US/09/529,063  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: PCT/JP98/04514  
; PRIOR FILING DATE: 1998-10-06  
; PRIOR APPLICATION NUMBER: JP 9-274674  
; PRIOR FILING DATE: 1997-10-07  
; NUMBER OF SEQ ID NOS: 117  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 55  
; LENGTH: 1109  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; US-10-414-378-55

Query Match 28 6%; Score 2612.5; DB 12; Length 1109;

Best Local Similarity 45.9%; Pred. No. 1.4e-224; Mismatches 343; Indels 59; Gaps 19;

Matches 517; Conservative 208; Mismatches 343; Indels 59; Gaps 19;

Qy 649 VLVEYPLRCLVLVAQVVAEMRRNGLSLISQVYFYQDVKREEMDYDKDIIMLQIGASLMD 708

Db 1 MLIEHPLRCLVLCAQVHAGMWRNGLFSLVNQIYVYHNVKCRREMFQDVVLTQVGSMD 60

Qy 709 PNKFLLLVQRYELAEFN----KTISTK--DQDLIKQYNTLIEEMLVQVLIYIGERVY 761

Db 61 PNHFLMILSRRELYQIFSTPDYGRFSEIHFHKDVQVQNNLTIEEMLYLIILVGERFS 120

Qy 762 PCVGNVTKREVTREIHLICIEPHPSAIAKNLPENNNENNGLENVINKVATFKKPGVS 821

Db 121 PGVQVNADEIRKRIIHLQISKPMHSELVKSIPEDENKETGMESVIEAVAHFKKGLT 180

Qy 822 GHGVYELKDSKDFNMVYHYSKTQHSKAEHMQRKRKQENKDEALPPPPPPFCFPAFS 881

Db 181 GRGMVELKPECAKEFNLYFHFSAEQSKAEAAQRKLRQNRDPTALPPVLPFCPLFA 240

Qy 882 KVINLLNCDIMMYILRTVFERAIDTDSNLWTEGMLQMAFHILALGLLEEKQOLKABEE 941

Db 241 SLVNILQSDVLMCTMGITLOWAVEHNGYAWSESMQLQRLVHLHIGMALQEEKHLENVTEH 300

Qy 942 -VTFDFYHKASRLGSSAMN---IQMLEKLGIPQLESGQDMITWILQFDTVKRLREKS 997

Db 301 VVTFFTQKISKPGAPKNSPSILAMLETQNAPLYEVHDMIRWILKTNVAVKMKRESS 360

Qy 998 -CLIVATTSGSEIKNDEITHDKAEKRRKAEARLHROKIMAOQSALQKNFTETKLM 1056

Db 361 PTPSPAETEGTIM---EESRDKDAERKAEARLRRKIMAOSEMQRHFIDENKEL 417

Qy 1057 YDNTEMPGKEDSIMEEESTPAVSYSRIALGKRGPSVTEKEVLTICILQOEQEVKIE 1116

Db 418 FOOTLELDASTSAVLDH--SPVASDMTLTALGPAQTVPEQRFVTCILQOEQEVKIES 475

Qy 1117 NAMVLSACVOKSTALTQHRKPIELSGEALDPLFMDPLAYGTGSCGHVMAHVCWQY 1176

Db 476 RAMVLAFAVQRTSVLSKNRSKFQ-DPEKYDPLFMHDPDLSCGTHSCGHIMHAHCWQY 534

Qy 1177 FEAVOLSSQO-----RIHVDLDESGEYLCPLKSLCNTVPIPILOPOKINSADAL 1231

Db 535 FDSVQAEQRQRLRLHTS-YDVENGFEPLCECLSNVPIPL-LSPRINFNNRLN-F 591

Qy 1232 AQLTLARWOTVLARISGYNIRHAKENPIPIFFNQGMDSTLEFHSILSFGVSESIKY 1291

Db 592 SDQPNLTQWIRTISQIKALQFLRKEESTPNNASTKNSENVDLQLPPEGFRDPKIPY 651

Qy 1292 SNSIKEMVILFATTYRIGLVKPPDRDPVRPMLTWSTCAFTQAIENLLGDEKGPLFGA 1351

Db 652 SESIKEMLTFTGTATYKVLGVHPNEEDPRVIMCWGSCAYTIQSIERILSDEKPLFGP 711



```
Db      319  NKLSSAENDRDLRREQLNVEK-----RQIM-----E 348
Qy      571  ECHAVMRCS---TSFSSKTVQSCGHSLETSYRVSDELVSIIH--LPLSRTLGLHV 625
Db      349  ECENLKLECSKLOPSAVKQSDTM-----TEKERILAQSAVSEVFRILQQAALSDAEN 399
Qy      626  RLRLGAVSRHLHEFVSFEDQVEVLVEYPLRCLVLVAQVVAEMWRNGLSLISVFFYQ- 684
Db      400  EIMRLSSLDNDSLAE--DNLKLMRIEVLKESKLSLQKEEL-----QMSLLKLNNEYEV 454
Qy      685  -----DVKCREMYDKDILMIQIGASLMDPNKFL---LLVLQRYEL---AEAFNKT 729
Db      455  IKSTATRDISLDELHD---LRLMLEAKEQLNDSISEKETLIAIEIELDRQNOEAKHM 511
Qy      730  ISTKQDQDLKQYN---TLBEMQLVLIYIVGERVYVPGVGNVTKREVTMRIHLLCIEPM 786
Db      512  ILIKDO--LSQONEGDSIISKQDL-----NDEKKRV-----HQLEDDKM 551
Qy      787  PHSAIAKNLPENENNETGLENVINKVATFKPGVGHGVVELKDESKLPNMYHYHYSKT 846
Db      552  D---ITKELDVQREKLQSEVALNDLHLTKQK-----LEDKVENLVD-----QLNKS 595
Qy      847  QHS-----KAEMOKKRRKQENKQDALPPPPPEFCPAFSKVINLLNCDIMMYILRTYFER 902
Db      596  QESNVSTQENLELKEHIRENEE-----LSRIRNEL-----MQ 629
Qy      903  AIDTDSNLTEGMLQAFHIALGLLEEKQOLKAPE--EVTDFYHKASRLGSSAMNI 960
Db      630  SLNODSNSPKDTLLKREAEVRLNKNLSELEQNLNKKVAFDVRMENEKVLVACEDV 689
Qy      961  QMLEK--LKGIPOLEGQKMITWLOMPD-----TVKRLREKSLIVATTSGSESI 1010
Db      690  RHOLEECLAGNOLSLKNTIVETLKMKEGIEALCWAKKRLLEAKYKTEIELSNA 749
Qy      1011  KN-----DEITH-----DKEAKRRKAE-----AARLHRQKIMQAM 1042
Db      750  RNLWTSALQLEHEHLIKLNQKDMKIEALKNIEQMDTDHKETKDVLSLSEEQKLTQL 809
Qy      1043  SALQKNFIEYHKMLNDYNTSEMPGK-----EDSTIMESES 1075
Db      810  INKKEIEFKERSSKLOELDYSQALRKNEILRTQIEBKDRSLGSMKEENHLQBEL 869
Qy      1076  TPAVSDYSRIA--IGPRKGPSVTE--KEVLTCILQEBEQEVKVIENNAMVLSACVOKSTAL 1131
Db      870  ERLREEQSRAPADPKTLDSTELASEVSQNLTIKEHLEBEIKHHQKIIEDQNGSKMQL 929
Qy      1132  TOHRGKPIELSGEALDPLFMDPDLAYGTYTSCGHHVHACWQKYFEAVQLSSQORIHVD 1191
Db      930  LQ-----SLQEQKKEDEBF-----RY-----QHEQMNATHQ 956
Qy      1192  LFDLESGEYLCPCKSLCNTVIPTI---IPLQPKINSENADALQALLTLARWIQTVLARI 1248
Db      957  LF--LEKDEEI---KSQKTIQELKTQLHEERQDIQDINSDFQE-----TKV 999
Qy      1249  SGYNIRHAKGENPIPIFNQMGDSTLEFHSILSGVSESSIKSYNSIKEMVILFATTIYR 1308
Db      1000  QSLNIENG-----EKHLSKAETERLVK-----1023
Qy      1309  IGLKVPDPDRPRVPMLTWSTCAFTIOAENLLGDEGKPLFGALQNR 1355
Db      1024  -GIR-----ERELEIKLLNEKNISLTQK-IDQLSKDEVGKLTQIITQOK 1064
```

## RESULT 5

```
US-09-815-242-5251
; Sequence 5251, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
```

```
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5251
; LENGTH: 996
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-5251
```

Query Match 2.0%; Score 180; DB 9; Length 996;

Best Local Similarity 19.4%; Pred. No. 2.8e-06;

Matches 223; Conservative 183; Mismatches 410; Indels 332; Gaps 55;

```
Qy      335  DFRQIFCQACLRPEEDSENPLIS-----RLMWDAKLY---KCARLILHELIFSFF 384
Db      20  DFSKI-----DNNELFLISGKTGSGKTMIFDAMTYALFGKASTQREENDLRSHF 69
Qy      385  MEMEYKFLFAMFV---KYRK--OLOKEYISDDHRSISITALSVMQFTVPTL---ARHLIE 438
Db      70  ADGKQPMSTVFEEQLNHRIVKVRQGYIKEGN-----TTKNAKDFVEMVDGKVEIR 123
Qy      439  EQNVISVTETLLEVLPEYLDNRNNFNFOGYQDKLGRVAV--ICDLKYLILSKPIWTE 497
Db      124  ESKVSG--TQFIELL-----GVNADQFRLQFILPQGEFKRFLISNREKQG 169
Qy      498  RLRMQF---LGFRSFLKILTCMQGMEIRQVGQHIIEVDPDWEAAIAIOMOLKNILLM 553
Db      170  ILRTLFDSEKFAIREILA-----EEVKKEKAO-----IENRYQIDLL 208
Qy      554  FOEWACADDE---LLLVAKECHKAVMRCSTSFSSKTVQSCGHSLET---KSYRVSE 607
Db      209  WQIEISFDDDKIKGLLEVATQIDKVIENIPL--LQARSKEILAFVNESKETAIKEYEIE 267
Qy      608  DLVSIHLPLSRITAG--LHVLSRLGAVSRHLHEFVSFEDQVEV--LVEYPLRCLVLVAQV 664
Db      268  -----KKTLENNILKDNINQLN--KNKIDFVLQKEQPEIEEIEAKLKLQDITNL 316
Qy      665  VAEWMRRNGLSL-----ISQV--FYQDVKCREEMDKDITMLQIGA-----SLMD 708
Db      317  LNYIENREKTIETKIASKKDISETNNKILNLDCKRNDKREKMLBENGDLISKISFID 376
Qy      709  PNKFLLLVLQRYELAEAFNKTISTKQDQLIKQVNTLIEEMQLVLIYIGERYVPGVNV 768
Db      377  KTRVLENDINKYQOQSVLNIERLFTDEQLDELNNLIKGL-----EKVEDSIGN-- 425
Qy      769  KEEVTMRIETHLLCIEPMPHSAIAKNLPENNETG-----LENN-----LEN 807
Db      426  -NESDYEKIIEI-----NNAITNINNEINVIKENEKAKDELKLLGSKQBLEN 472
Qy      808  VLNKVATFKKPGVGHGVVELKDESKLPNMYHYHYSKTQ---HSAEKHMOKRRKQENK 864
```



Db 473 QINEEKTI-----LKNLEIKLDYDKSKLDLNDKSFISEIKSAVKIG 515  
QY 865 DEALPPPPPPFPCAFSKVINLLNCDIMMYILRTVFERAIDTDSNLTGMLQAFHILA 924  
Db 516 DQ-----CPICGNEIQDLGHDFDSTAKRONEIKEIEANIHMTES--NIAVHSE 564  
QY 925 LGLLEEK-QOLQKAPEEVTFDYHKASRLGSSAMNIQMLEKLGIPQLEGQKDMITWI 983  
Db 565 IKFVNEKISINIKTQSDLSLEVLNKRLLLENALNQDLNKF--IEQMKEDKDLT-- 620  
QY 984 LQMFDTVKRLRE-----KSC--LIV--ATTSGSESKNDEITHDKKAEKRKAARLH 1034  
Db 621 LOIHNKQLRNLKNESELKICRNLITFETLSKYNNITNFEVDYKVIQDVNQHOEHSNQI 680  
QY 1035 ROKIMAQMSALQKNFIETHKL-----MYDNTSEMPGKEDSIMEEESTPAVSD----- 1081  
Db 681 EDKLI-QLS--QRKLIQONNLNHYENQLETYNNDLEL--NEQSIEMEMSRNLNTDNDIN 735  
QY 1082 -----YSRIALGPKRGPSVT-EKEVLTCTILCOEEQEVK----- 1113  
Db 736 ELIAHRGEELEQKRDYTKRYHEFEFEMIEARLESITKDELDSKLDKDDYELKKGKMN 795  
QY 1114 --IENNAMVLSAC---VOKSTALTOH-----RGRPIELSGE 1144  
Db 796 TLIDEVSAVHYQCONNIKTQSTIVSHINQELKDQOEIQLAEITLSCGKNNKNTLENF 855  
QY 1145 A-----LDPLFMPDPLAIGTYTSCGHVMHAWCQKYEAVQ--LSSQQRIHVDLFDLESG 1198  
Db 856 VLIYLDIQIAQANRLATMSDNRYQLIRR-----EAVSHGLSG---LEIDVFDLHNS 905  
QY 1199 EYLCPLCKSLCNTVPIIPLQPKINS-ENADALAOQLLLTLARWIOVRLARISYNIHAK 1257  
Db 906 -----KSRHISLSGGETFQSSALALGSLSIQVQOQSG----- 938  
QY 1258 GENPIPIFNQMGDSTLEFHSILSGVESSIKYSIKEMVTLFATTYRIGLKVPDPPE 1317  
Db 939 GISLESIFIDEGFG--TLDOETL-----ETVLDITLLNKAETGRWGIISHVSE 984  
QY 1318 RDRPRVPL 1325  
Db 985 LKNRIPLV 992

RESULT 6  
US-09-815-242-12141  
; Sequence 12141, Application US/09815242  
; Patent No. US2002061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12141  
; LENGTH: 1009  
; TYPE: PRF  
; ORGANISM: Staphylococcus aureus  
US-09-815-242-12141

Query Match 2.0%; Score 180; DB 9; Length 1009;  
Best Local Similarity 19.4%; Pred. No. 2.9e-06;  
Matches 223; Conservative 183; Mismatches 410; Indels 332; Gaps 55;  
QY 335 DFRIFCOACUREPDSNPCLIS-----RLMLDAKLY-----KGARKILHELIFSSFF 384  
Db 20 DFSKI-----DNNELFLISGKTGKTMIFDAMTYALFGKASTEQREENDLRSHF 69  
QY 385 MEMEYKKLFAMEFV---KYYK-QLQKEYISDDHRSISITALSVQMFVPTL--ARHLIE 438  
Db 70 ADGQKPSVTFEFOLNHRHYKVRQGPYKEGN-----TTKNAKFDVFEMVDGKYEIR 123  
QY 439 EONVISVITETLLEVLPEYLDNRNKNFQGYSDKLGRIYAV-ICDLKYLILSKPTIWE 497  
Db 124 ESKVISG-TQPIIEL-----GVNADQFROLFILPQGEFRFLISNREKOG 169  
QY 498 RLRMQF----LEGFRSEFKILTCMQGMEIEIRROVGHIEVDPDEAAIAIOMQLKNILM 553  
Db 170 ILRTLDFSEKFEATREILK-----EEVKKEKAQ-----IENRYQOJDL 208  
QY 554 FQEWACDEE---LLLVAYKECHKAVMRCSTSFSSSKTVVQSGCHSLET--KSYRVSE 607  
Db 209 WOETIESFDDDKIKGLLEVATQIDKVIENIPL-LQARSKEILAFVNESKETAKEYEITE 267  
QY 608 DLVSIHLPLSRTLAG--LHVLSRLGAVSRHLHEFVSFDFQEV-LVEYPLRCLVLVAQV 664  
Db 268 -----KRTLENNILKDNINQLN--KNKIDFVLKQEQPEIEIEAKLLQDITNL 316  
QY 665 VAEWRRNGLSL-----ISQV-FYYQDYKCREMYDKDIIIMIQIGA-----SLMD 708  
Db 317 LNYIENREKIEITKIAKSKKDISETNNKTLNLDCKRNDIDKEKKMLEENGDLIESKIFD 376  
QY 709 PNKFLLLVQRYELAAEPNKTISTKDDOLIKOYNTLIBEMLQVLIYVIGERYVPGVNT 768  
Db 377 KTRVLFNDINKYQOXYLNIERLRTDEOLGDELNNLIKGL-----EKVEDSIGN-- 425  
QY 769 KEVTMPREIITHLLCIEPMPHSAIAKNLPENNETG-----LENN--LEN 807  
Db 426 -NESDYEKIIEI-----NNAITNINNEINVIKENEKAKDELKLLSGKQELN 472  
QY 808 VINKVATFKPGVSGHGVYELKDESLDFNMFYHYHYSKTO---HSAEHEMOKRRKQENK 864  
Db 473 QINEEKTI-----LKNLEIKLDYDKSKLDLNDKSFISEIKSAVKIG 515  
QY 865 DEALPPPPPPFPCAFSKVINLLNCDIMMYILRTVFERAIDTDSNLTGMLQAFHILA 924  
Db 516 DQ-----CPICGNEIQDLGHDFDSTAKRONEIKEIEANIHMTES--NIAVHSE 564  
QY 925 LGLLEEK-QOLQKAPEEVTFDYHKASRLGSSAMNIQMLEKLGIPQLEGQKDMITWI 983  
Db 565 IKFVNEKISINIKTQSDLSLEVLNKRLLLENALNQDLNKF--IEQMKEDKDLT-- 620  
QY 984 LQMFDTVKRLRE-----KSC--LIV--ATTSGSESKNDEITHDKKAEKRKAARLH 1034  
Db 621 LOIHNKQLRNLKNESELKICRNLITFETLSKYNNITNFEVDYKVIQDVNQHOEHSNQI 680  
QY 1035 ROKIMAQMSALQKNFIETHKL-----MYDNTSEMPGKEDSIMEEESTPAVSD----- 1081  
Db 681 EDKLI-QLS--QRKLIQONNLNHYENQLETYNNDLEL--NEQSIEMEMSRNLNTDNDIN 735  
QY 1082 -----YSRIALGPKRGPSVT-EKEVLTCTILCOEEQEVK----- 1113  
Db 736 ELIAHRGEELEQKRDYTKRYHEFEFEMIEARLESITKDELDSKLDKDDYELKKGKMN 795



```
; APPLICANT: Glatt, Karen
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Hoersh, Sebastian
; TITLE OF INVENTION: IDENTIFICATION, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
; TITLE OF INVENTION: OF CERVICAL CANCER
; FILE REFERENCE: MRI-035
; CURRENT APPLICATION NUMBER: US/10/171,311
; CURRENT FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: US 60/298,159
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,155
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/335,936
; PRIOR FILING DATE: 2001-11-14
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 3899
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-171-311-4

Query Match          1.8%   Score 168.5; DB 15; Length 3899;
Best Local Similarity 18.7%; Pred. No. 0.00034;
Matches 201; Conservative 165; Mismatches 364; Indels 347; Gaps 50;

Qy      208 EMTWEEKEPELPLOIREKNERYICVLFNDEHHSYDRHVYSIQRLADCELAEAQLHTTA 267
        ||| : || :|| : | : | : | : | : | : | : | : | : | : | : | : | :
Db      716 EMTL-QINELQREIELROEEK-----EKTLEQEVEQLQ--LKTELLEKOMK--- 760

Qy      268 IDEGRRAVAGAYAAOAEKEDIKSHENVSQHPLHVEVLHSEIMAHQFALRGLGSMWN 327
        ||| : || :|| : | : | : | : | : | : | : | : | : | : | : | : | :
Db      761 -EKENDLQEFQAOLAEANSILKDKKTKLEDMLK-----IHTPVSOBERLI----FLD 807

Qy      328 KIMSYSDF---ROIQCACLREPDPSENCLISRLMDAKLYKGARKILHELIFSFFF 384
        ||| : || :|| : | : | : | : | : | : | : | : | : | : | : | : | :
Db      808 SIKSKSDSWWEKEI---EIIEENEDLKQCI-----QLNEETEKORNTSFSAEKN 856

Qy      385 MEEMEYKKLFAMFEVKYKQLOKEY-----ISDDHDRSITALSVMQFTPTTARHLIEE 439
        ||| : || :|| : | : | : | : | : | : | : | : | : | : | : | : | :
Db      857 FEVN-----QOELEEYACLLKYKDDLEDSKN-----KQLEY 889

Qy      440 QNVISVITEITLVLPBYLDNN---KFNFQGYSDQKLGRVAVICDLAYILISKPTIW 495
        ||| : || :|| : | : | : | : | : | : | : | : | : | : | : | : | :
Db      890 KSKLKALNEEL-----HLQRINPITVKMKSSVPDEDK--TFVAETLEMGEVVEKDTTEL 941

Qy      496 TERLRWFLEGFRSFKILLTCMQMGEIRIQVGQ-HIEVDPDWEFAATAIQMLKNILMF 554
        ||| : || :|| : | : | : | : | : | : | : | : | : | : | : | : | :
Db      942 MEKLEVTTREK-----LELSQRSLSDELQKKGHEISFLNBEVKSLSQKEQVSLRC 994

Qy      555 QEW-----CACDEE-----LLLVAYPECHKAVMCRSTFSISSKTVIVSCG 595
        ||| : || :|| : | : | : | : | : | : | : | : | : | : | : | : | :
Db      995 RELEIIHNHNAENVQSCDTQVSSLDGVTMTSRGAEGSVSKYNKSGFEESKIMVED-- 1052

Qy      596 HSLETksYRVSED-----LWSIHPLRSRTLGLH-VRLSRIGAVSRHFEVFDFQVEV 649
        ||| : || :|| : | : | : | : | : | : | : | : | : | : | : | : | :
Db      1053 -KVSNFMNTVGEESKQQLIILDHLPSTKESSLRATQPSENDKLOKELNLVKSQNDLRL 1111

Qy      650 LVEYPLRCLVLVIAQVVAEMRRNCGLSISQVYYQDVKCREMYDKDIIMLQTGASLM DP 709
        ||| : || :|| : | : | : | : | : | : | : | : | : | : | : | : | :
Db      1112 QMEAQRICLSLVVS-----THVDQVREYME-----NEKDALKCSLKEELIFA 1153

Qy      710 NKFLLLVLQR-YELAAFNKTIstkOD-----LIKQYNTLIEB---MLQVLIYVGER 759
        ||| : || :|| : | : | : | : | : | : | : | : | : | : | : | : | :
Db      1154 OEKIKEQLKHOLELTQMTQTDGSGKPLHLIGLQKAVSECSEYFLOTLCVSLGEY 1213

Qy      760 YVPGVGNVTKEEYTMREIIHLCTEPMPHSAIAKLNLENNETGTLENINKVATFKKP 819
        ||| : || :|| : | : | : | : | : | : | : | : | : | : | : | : | :
Db      1214 YTPAL----KCEV-----NAEDKEN----- 1229

Qy      820 VSGHGVELKDSESLKDFNMYYFYHYSKTQHSAEHOMKQRKQENKDEALPPPPPFCPA 879
        ||| : || :|| : | : | : | : | : | : | : | : | : | : | : | : | :
```

```
Db 808 SIKSKSDVWEKEI--EILIEENEDLKQOCI-----QLNEEIEKQRTFFFAKN 856
Qy 385 MEMEYKFLFAMEFYKYKQLOKEY-----ISDDHRSISITALSVMFTVPTLARHLIE 439
Db 857 FEVN-----YQLOEYACLLKVKDDLEDSKN-----KOELEY 889
Qy 440 QNVISVITETLLEVPDYDRNN---KFNQGYSDKLGRRYAVICDLKYLILSKPTIW 495
Db 890 KSKLALNEEL-----HLQRINPTTVKMKSSVFDEDK--TFVAETLEMGEVVEKDTTEL 941
Qy 496 TERLRMQFLGFRGFLKILTCMOGMEIRROVGO--HIEVDPDWEAAIAIOMQLKNILMF 554
Db 942 MEKLEVTKREK-----LELSQRLSDLSQKQKHGEISFLNEEVKSLKQKEQVSLRC 994
Qy 555 QEW-----CACDEE-----LLLVAYKECHKVAMRCSTSFSSSKTVVQSCG 595
Db 995 RELEIIINHNRAENVQSCDTQVSSLLDGVVMTSRGAEVSKVKNKSGEESKINVED-- 1052
Qy 596 HSLKTSKYRVSED-----LVSIHLPLSRTLGLH--VRLSLRGAVSRHLHEFVSFDFQVEV 649
Db 1053 -KVSFENNTVGEESKQEOILDLHLPSTVKESLRATOPSENDKLOKELNVKSEQNDLRL 1111
Qy 650 LVEYPLRCLVLVAQVVAEMRRNGLSLSQVFFYQDVVKCREMYDKDIIMLQIGASLMDP 709
Db 1112 QMEARICLSLVYS-----THVDQVREYME-----NEKDALCSLKEELIFA 1153
Qy 710 NKFLLLVLQOR--YELAEAFNKTISTKQDQ-----LIKQYNTLIEE-----MLQVLIYVGER 759
Db 1154 QEEKIKELQKIHOLELQTMKTQETGCKPLHLLIGLKQKAVSECSYFLOTLCVSLGEY 1213
Qy 760 YVPGVGNVTKEEVTMREIHLCLTEPMPSHAIAKNLPENNETGLENVINKVATFKKPG 819
Db 1214 YTPAL-----KCEV-----NAEDKEN-----1229
Qy 820 VSGHGVYELKDESILKDFNMVYHYHYSKTOHSAEHMOKRRKQENKDEALPPPPPEFCPA 879
Db 1230 -SGDYISENEDPELODYRYEVODFOENMHT-----LLNKVTEYNK-----1269
Qy 880 FSKVINLLNCIDIMYILRTVPERAIDTDSNLW---TEGMLQAFHIALGLLEEKQOLQK 936
Db 1270 -----LLVLQTRL-----SKIWQOTDGM-----KLEFGE-----N 1296
Qy 937 APEEVEVTF-----DFYHKASRLGSSAMNIQMLEKLGIPOLEGQKDMITWI 983
Db 1297 LPKEETEFLSIHSQMTNLEDIVNHK-----SKLSSLDLEKTKLEEQVQVLESLSL 1350
Qy 984 LQMFDTVKRLREKS--CL-----IVATTSGSESINKDEITHDKKAERKRAEARLHRQ 1036
Db 1351 QOOLKETQNYEAETHCLOKRLQAVSESTVPPSLPVDVSVITESDA-----Q 1397
Qy 1037 KIMQMSALQKNFIETHKLMVDNTESEMPGKEDSIMEEESTPAVSDYSRIALGPKRGPSVT 1096
Db 1398 RTMYPGCVKNNI-----DGTIEFGS--REGVKEETNI-----VKLL 1432
Qy 1097 EKEVLTICLOQEQEVKIENNAMVLISACVQKSTALTOHRGKPIELSGEALDPLFMDPDIA 1156
Db 1433 EKQY-----QEOLEEEVAKVIVSMSIAFAQOTELSR-----ISG-----1466
Qy 1157 YGTYTGSGCHVHVAWCWKQ--YFEAVQLSSQRIHVDLF---DLSGEYLCPLCKSL 1208
Db 1467 -GKENTASSKQAHAVCOEQEYFNEMKL--SQDQIGFQTFETVDVVKFEKFKPLSKEL 1521
```

## RESULT 10

```
US-10-171-311-8
; Sequence 8, Application US/10171311
; Publication No. us20030087270A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Chen, Yan
; APPLICANT: Zhao, Xumel
; APPLICANT: Monahan, John
```

```
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Glatt, Karen
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Hoersh, Sebastian
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
; TITLE OF INVENTION: OF CERVICAL CANCER
; FILE REFERENCE: MRI-035
; CURRENT APPLICATION NUMBER: US/10/171,311
; CURRENT FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: US 60/298,159
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,155
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/335,936
; PRIOR FILING DATE: 2001-11-14
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 3917
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-171-311-8

Query Match 1.8%; Score 168.5; DB 15; Length 3917;
Best Local Similarity 18.7%; Pred. No. 0.00034;
Matches 201; Conservative 165; Mismatches 364; Indels 347; Gaps 50;

Qy 208 EMTINEEKELPPELOIREKNERYCVLNFNDEHSHDVHVIYSQRLALDCELAQAOLHTTA 267
Db 716 EMTL--QINELQKEIILROEEK-----EKGTEQEVOELQ--LKTELLEKQMK--- 760
Qy 268 IDEKRRAYKAGAYAAQCAEKEDIKSHSENVSQHPLHVEVLHSEIMAHQFALRLGSMWN 327
Db 761 -EKENDLOEFAQLEAENSLDKKTKLEDMLK-----IHPVSOEERLI-----FLD 807
Qy 328 KIMSYSSDF---RQIFQACLRPEPSENPCLSRLMLMDAKYLGARKLHELILFSSFF 384
Db 808 SIKSKSDVWEKEI--EILIEENEDLKQOCI-----OLNEETEKQRTFFFAKN 856
Qy 385 MEMEYKFLFAMEFYKYKQLOKEY-----ISDDHRSISITALSVMFTVPTLARHLIE 439
Db 857 FEVN-----YQLOEYACLLKVKDDLEDSKN-----KOELEY 889
Qy 440 QNVISVITETLLEVPDYDRNN---KFNQGYSDKLGRRYAVICDLKYLILSKPTIW 495
Db 890 KSKLALNEEL-----HLQRINPTTVKMKSSVFDEDK--TFVAETLEMGEVVEKDTTEL 941
Qy 496 TERLRMQFLGFRGFLKILTCMOGMEIRROVGO--HIEVDPDWEAAIAIOMQLKNILMF 554
Db 942 MEKLEVTKREK-----LELSQRLSDLSQKQKHGEISFLNEEVKSLKQKEQVSLRC 994
Qy 555 QEW-----CACDEE-----LLLVAYKECHKVAMRCSTSFSSSKTVVQSCG 595
Db 995 RELEIIINHNRAENVQSCDTQVSSLLDGVVMTSRGAEVSKVKNKSGEESKINVED-- 1052
Qy 596 HSLKTSKYRVSED-----LVSIHLPLSRTLGLH--VRLSLRGAVSRHLHEFVSFDFQVEV 649
Db 1053 -KVSFENNTVGEESKQEOILDLHLPSTVKESLRATOPSENDKLOKELNVKSEQNDLRL 1111
Qy 650 LVEYPLRCLVLVAQVVAEMRRNGLSLSQVFFYQDVVKCREMYDKDIIMLQIGASLMDP 709
Db 1112 QMEARICLSLVYS-----THVDQVREYME-----NEKDALCSLKEELIFA 1153
Qy 710 NKFLLLVLQOR--YELAEAFNKTISTKQDQ-----LIKQYNTLIEE-----MLQVLIYVGER 759
Db 1154 QEEKIKELQKIHOLELQTMKTQETGCKPLHLLIGLKQKAVSECSYFLOTLCVSLGEY 1213
Qy 760 YVPGVGNVTKEEVTMREIHLCLTEPMPSHAIAKNLPENNETGLENVINKVATFKKPG 819
Db 1214 YTPAL-----KCEV-----NAEDKEN-----1229
Qy 820 VSGHGVYELKDESILKDFNMVYHYHYSKTOHSAEHMOKRRKQENKDEALPPPPPEFCPA 879
```

```
Db 1230 -SGDYISENEDPELDQRYRYEVQDFQENMHT-----LLNKYTEYNK----- 1269
QY 880 FSKVINLLNCDIMYILRTVFERAIDTDSNLW---TEGMLQMAFHILALGLLEEKQOOLQK 936
Db 1270 -----LLVLQTRL-----SKIWOOTDGM-----KLEFEGEE-----N 1296
QY 937 APEEVTF-----DFYHKASRLGSSAMNIQMLLEKLGIPQLEGOKDMITWI 983
Db 1297 LPKETEFLSHSQMTNLEDIDVNHK-----SKLSSQDLKTKLEEQVQLESLSL 1350
QY 984 LQMFDTVKRLREKS--CL-----IVATTSGSESINKDEITHDKKAERKRAEAAARLHRQ 1036
Db 1351 QOOLKETEQNYEAEIHCQLKRLQAVSESTVPPSLPVDVSVITESDA-----Q 1397
QY 1037 KIMAQMSALQKNFIETHKLMYDNTSEMPGKEDSIMEESTPAVDYSRIALGPKRGPSVT 1096
Db 1398 RTMYPGSCVKKNI-----DGTIEFSG-EFGVKEETNI-----VKLL 1432
QY 1097 EKEVLTCILCOEQEVEKVIENNAWVLSACVOKSTALTOHRGKPIELSGEALDPLFMDPDLA 1156
Db 1433 EKQY-----QEULEEVAKVIVMSIAFAQOQTELSR-----ISG----- 1466
QY 1157 YGTYTSCGHVMAHVCWK--YFEAVQLSSQORIHVDLF---DLESGEVLCPLCKSL 1208
Db 1467 -GKENTASSKQAHAVCQEQHGFENMKL-SQDQIGFOTFETVDVKFEKPELPSKEL 1521
```

## RESULT 11

```
US-10-171-311-6
; Sequence 6, Application US/10171311
; Publication No. US20030087270A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Chen, Yan
; APPLICANT: Zhao, Xumei
; APPLICANT: Monahan, John
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Glatt, Karen
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Hoerish, Sebastian
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
; TITLE OF INVENTION: OF CERVICAL CANCER
; FILE REFERENCE: MRI-035
; CURRENT APPLICATION NUMBER: US/10/171,311
; CURRENT FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: US 60/298,159
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,155
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/335,936
; PRIOR FILING DATE: 2001-11-14
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 3925
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-171-311-6
```

```
Query Match 1.8%; Score 168.5; DB 15; Length 3925;
Best Local Similarity 18.7%; Pred. No. 0.0034;
Matches 201; Conservative 165; Mismatches 364; Indels 347; Gaps 50;

QY 208 BMTWEEKELPPELQIREKNERYCVLFNDEHHSHVYISLQRLDCELAEALHTTA 267
Db 716 EMTL-QINLEKQETIELRQEK-----EKGLEGEVQBLQ-LKTELLEKQMK--- 760
QY 268 IDKEGRRAVKAGACQAEAKEDIKSHSENYSOHPLHVEVLHSEITMAHQKFAIRLGSMWN 327
Db 761 -EKENDLQERQAFQAEANSILKDEKKTLEDMLK-----IHTPVQSEERLI-----FLD 807
```

## RESULT 12

```
US-10-205-219-169
; Sequence 169, Application US/10205219
; Publication No. US20030138803A1
; GENERAL INFORMATION:
; APPLICANT: Warner-Lambert Company
; APPLICANT: Lee, Kevin
; APPLICANT: Dixon, Alistair
```

```
QY 328 KIMSYSDF---RQIFCOACLREBPDSNPCLISRLMLWDAKYLGARKILHELIFSFF 384
Db 808 SIKSKSDSWEREI--EILIEENEDLKQOCI-----OLNEETEKQRTSFAEKN 856
QY 385 MEMEYKFLFAMEFVKYKQLOKEY-----ISDDHRSITATSALSVQMFVPTTLARHLIEE 439
Db 857 FEVN-----YQLEQEEVACLKVKDDLEDSEN-----KQLELEY 889
QY 440 QNVISVITELLEVLEPEYLDNN---KPNFQGSQDKLGRVYAVICOLKYLISKPTIW 495
Db 890 KSKLKALNEEL-----HIGRIINPTTVKMKSSVFEDDK--TFVAETLEMGEVVEKDTTEL 941
QY 496 TERLRQFLGRFSFLKILTCMQGMEERQVGO-HIEVDPDWGAATAIOMQLNILLMF 554
Db 942 MEKLEVTREK-----LELSQRLSDLSQKQKHGEISFLNEEVSFKQEKQVSLR 994
QY 555 QEW-----CACDEE-----LLIVAYKECHKAVMRCSTFISSSKTVVQSCG 595
Db 995 RELEIINHNAENAVQSCDTQVSSLLDGVVTMTSRGAEGSVSKYNKSGEESKIMVED-- 1052
QY 596 HSLKETSYRVED-----LVSIHPLSRTLAGLH-VRLSRLGAVSRHLHEFVSFEDFOVEY 649
Db 1053 -KVSFENMTVGESKQBLILDHLPSTKESLRATQPSENDKLOKELNVLKSEQNDURL 1111
QY 650 LVEYPLRCLVLVAQVVAEMRRNRLSLISQVFFYQDYKCREEMDYDKDIIMLQIGASLMDP 709
Db 1112 QMEAQRICLSLVYS-----THVDQVREYME-----NEKDALKSLKEELIFA 1153
QY 710 NKFLLLVLQR-YELAEAFNKTISTKODD-----LIKQYNTLIEE---MLQVLIVIGER 759
Db 1154 QEEKIKELQIHQLELOTKMTQETGDEGKPLHLILGLKQAVSECSYFLOTLSVIGEY 1213
QY 760 YVPGVGNVTKEEVTMREIHLCLIEPMHSAIAKNLPENENNETGLENVINKVATFKPKG 819
Db 1214 YTPAL---KCEV-----NAEDKEN----- 1229
QY 820 VSGHGVVELKDESILKDFENFYHYKTOHSAEHMQRKRQENKDEALPPPPPEFCPA 879
Db 1230 -SGDYISENEDPELDQRYRYEVQDFQENMHT---LLNKYTEYNK----- 1269
QY 880 FSKVINLLNCDIMYILRTVFERAIDTDSNLW---TEGMLQMAFHILALGLLEEKQOOLQK 936
Db 1270 -----LLVLQTRL-----SKIWOOTDGM-----KLEFEGEE-----N 1296
QY 937 APEEVTF-----DFYHKASRLGSSAMNIQMLLEKLGIPQLEGOKDMITWI 983
Db 1297 LPKETEFLSHSQMTNLEDIDVNHK-----SKLSSQDLKTKLEEQVQLESLSL 1350
QY 984 LQMFDTVKRLREKS--CL-----IVATTSGSESINKDEITHDKKAERKRAEAAARLHRQ 1036
Db 1351 QOOLKETEQNYEAEIHCQLKRLQAVSESTVPPSLPVDVSVITESDA-----Q 1397
QY 1037 KIMAQMSALQKNFIETHKLMYDNTSEMPGKEDSIMEESTPAVDYSRIALGPKRGPSVT 1096
Db 1398 RTMYPGSCVKKNI-----DGTIEFSG-EFGVKEETNI-----VKLL 1432
QY 1097 EKEVLTCILCOEQEVEKVIENNAWVLSACVOKSTALTOHRGKPIELSGEALDPLFMDPDLA 1156
Db 1433 EKQY-----QEULEEVAKVIVMSIAFAQOQTELSR-----ISG----- 1466
QY 1157 YGTYTSCGHVMAHVCWK--YFEAVQLSSQORIHVDLF---DLESGEVLCPLCKSL 1208
Db 1467 -GKENTASSKQAHAVCQEQHGFENMKL-SQDQIGFOTFETVDVKFEKPELPSKEL 1521
```

```
; APPLICANT: Brooksbank, Robert
; APPLICANT: Pinnock, Robert
; TITLE OF INVENTION: Identification and Use of Molecules Implicated in Pain
; FILE REFERENCE: WL-A-018200
; CURRENT APPLICATION NUMBER: US/10/205,219
; CURRENT FILING DATE: 2002-07-24
; PRIOR APPLICATION NUMBER: GB 0118354.0
; PRIOR FILING DATE: 2001-07-27
; NUMBER OF SEQ ID NOS: 197
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 169
; LENGTH: 2649
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Pemphigoid antigen
US-10-205-219-169

Query Match 1.88; Score 165; DB 12; Length 2649;
Best Local Similarity 17.8%; Pred. No. 0.00035;
Matches 349; Conservative 262; Mismatches 644; Indels 710; Gaps 92;

QY 18 EISAELPQPORLASWMDQOVDFYTAFLHLAOLVPEIYFAEMDPD---LEKQEEVSOMS 74
DB 417 ELMRELDQKEENIKS--VQEIARQLLEHNPARTTEAYRAAMQOTOWSWILQOCQVEQH 474
QY 75 IFTPLEWYLFGEPPDICLEKLKH-SGAFOQLCGRVFKSGEFTYSC-RDCAIDPTCVLCMDC 132
DB 475 IKENTAYFEFFNDAKEATDYLRNLKDAIQ-----RKYSRDSRSSIHK-----LEDL 520
QY 133 FQDSVHKHRYKMHSTGGGFCDCGDTAEAWKTGFCVNHPEPGRAGTI-----KNSRCPLN 188
DB 521 VQESMEKEKELQYKST-----IANLMGKAKTIQIKRPNSDCPLK 561
QY 189 EEVIVQARKTFPSVIKY-VVEMTWEEKEELPELQIREKNERYCVLFNDEHHSYDHVI 247
DB 562 TSIPIKA-----ICDYRQIEITIKYKDE-----CVLANNSHRAKWKVI 599
QY 248 YSLQALDCBLAEALQHTT--AIDKEGRRAVAKAGAAQCAQEDIKSHSENVSQ--HPL 303
DB 600 SPTGN-----EAMVPSVCFTVPPNKEAV-----DLANRIEQOYQNVLTWHES 643
QY 304 HVEV-----LHSET-----MAHQKFALRLGSGWNNKIMSYSDPRQI 339
DB 644 HINNKSVSWHYLINEIDRIRASNVASIKTMLPGEHQOVLNSLQSRPEFLEDSQE-SQV 702
QY 340 FCOACLEEDPDSENPLISRLMLWDAKLYKGARKILHELIFSSFFMEME----- 388
DB 703 FSGSDITQLEKEVNVK-----KQYQOELLKSAEREQEEVSNLYISVNRNIRLNCED 758
QY 389 -----YKKLFAM-EFVKYKQLOKEYISDDHRSISITALSVMF----- 427
DB 759 RLRIQRTPLERDLHESVFRITEQELKKELER--LKDDLQ---TITNKEEFFSQAAA 813
QY 428 --TVPTLARHLIEQNVISVITETLLVLPYLDRNNKFNK---QGYSDQKLGRVY---- 478
DB 814 SSSVPTLR-----SELNVVLQMNQVISMSTYIDKLKTVNLVKNTOAAEALVLYETKL 869
QY 479 -----AVICDLKYI--LISKPTIWTREL----- 499
DB 870 CEEAVIADKNNIENLISTLKQMRSEYDEKRVFHALELQKAKAISDEMFTYKERDL 929
QY 500 -----RMOFLEGRFSFLK-----ILTCMOGMEETRRQ 526
DB 930 DFDWHEKADQLVERQNVHVDNRLRDLLEGKSLKYRYDTHPLDDWIQQVETTORQ 989
QY 527 VGOHIEVDPPWEAAIAIQMKNLILLMFQEWCAQDEELLVAYKECHKAVMRCSTSFSS 586
DB 990 IQEN--Q-SPENSKTATQLAQKQMLV-----SEIEMKQSKMDECCQKAEQYSA----- 1034
QY 587 SKTVQSCGSHLETYSRV-----SEDLVSIHLPLSRTLGLHVLRLSR 629
DB 1035 --TVKD---YELQTMITYRAMVDVSDQKSPVKRRRRMQSSADLI-----IQEFMDLRTRYTA 1083
```

```
QY 630 LGAVSRHLHEFVSPEDFQVEVLVEYPL-RLVLVAQVVAEMRRNGLSLISQVFFYQDYKVC 688
DB 1084 L--VTLMQTVIKFAGDSLKRLEEIEIKR-----KETSHEGAYSDQLQ 1124
QY 689 REEMYDKDIIMLQIGASIMDPNKF-----LLVLQRYELAAEFNKTISTKQDDL 737
DB 1125 RQK-----ATVLENSKLTGKISELERMVAELKKQKSRVEELPKVRAENEL 1172
QY 738 IKQYNTLIEMLQ-VLIYIVGERYVPGVGNVTK-EVTMRI--IHLCLIEPMPHSAIAK 793
DB 1173 RKQORNVEDISLQIRAESEAKQYRRELETIVREKEAARELERVROUTIE-----AEAK 1227
QY 794 NLPENENNETGLENVINKVA--TFKKPGVSGHGVYELKDESILKDPENMYFYHYSKTHSKA 851
DB 1228 RAAVEEN---LLNFRNQLEENTFTRTLEDH--LKRKDLSLNDL-----EOKN 1271
QY 852 EHMOKKRRKOENKDEALPPPPPPCFAPFASVINVLLNCIDIMMYILTRTFERAIDTDSNLW 911
DB 1272 KLMEELRRKRDNHEEL-----KLIKOMEKDL-----AFQKV----- 1304
QY 912 TEGMLQMAFHILALGLLEEKQOOLQKAPEEVT-FDFYHKASRLGSSAMNIOMLEKLKGI 970
DB 1305 -----AEQOLKEKQKIELEARRKITEIQTCTRENALPVCPTQATSCRAVTGL 1352
QY 971 POLBQKDMITWILOMFDTVRLREKSLIVATTSGSESIKND--EITHDKB--KAERKR 1026
DB 1353 QQ-BHDKQKABELKQOVDEL-----TAANKRKAQDMRELTYELNALOLEKTS 1398
QY 1027 KAAEARLHROKI-----MAQMSALQKNFIETHKLIMYDNTSEMPCKEDSIMEE 1073
DB 1399 SEEKARLLKDKDLTNNTLRCLKLELERKQAEKGYSOQLRELGRQLNQTGKAEAMQE 1458
QY 1074 ESTPAVSYSRIALGPKRGPSTVEKVLTCILCOEQEVEKVIENNANVLISACVQKSTALTQ 1133
DB 1459 -----ASDLKI-----KRYQL-ELESN-----HEKGLQREVDRITRAHVAEKN--IQ 1502
QY 1134 HRGPIELSGBALDPLFMDPDLAYGTGTGSGHVMHVCWKYFPAVOLSQORTIHDVLF 1193
DB 1503 HLNSQI-----HSPRDEKELERLQICORKSOHLREQ 1533
QY 1194 DLESGEYLCPLCKSLCNTVPIIPILOPQKINSENADALAQLTLARWTQTVLARISGYNI 1253
DB 1534 FEKSHQQL-----LQNIKAENKNDKIQRL----- 1558
QY 1254 RHAKGENPIPIFFNOGMDSTLEPHSILSGFVSESIKYNSIK-----EMVILFA 1303
DB 1559 -----NEELEKSN-ECAEMLKQKVEELTRQNNETKLMQRIQAESENIULEK 1604
QY 1304 TTIYRIGLKVPPDERDPRVPMLTWSTC-AFTIQAIENLLGDGKPLFGALQNRQINGLKA 1362
DB 1605 QTIQQ-----RCEALKTQA-----DGFQDKQLASTNEHLHKQTKT 1638
QY 1363 LMQP-----AVAQRI-----CPQ--VLIQ--KHLVRLSVLVLPNIKSEDTPC 1401
DB 1639 EQDFQRTKCLEEDLAKSONLVSEPKQKQDOQNIITQNTKKEVRNLNAELNASKEE----- 1694
QY 1402 LLSIDLPHVLVGAVLAFPLSLYWDVDPDLQPSVSSSYNHYL-----FHLIT-----MAHML 1453
DB 1695 -----KRRCEQKVOLOAQVOQVQLNNLKKVQDELHLKLTIEEQWTHRK 1736
QY 1454 QILLTVDTGRLPLAQVOEDESEAHSSAFFAEISQYTSIGCIDPGWYL-WYSLANGITTP 1512
DB 1737 MVLQEEESG---KFKQSAEEFRKKMEKLMESKVITEN-----DISGIRLDFVSLQ----- 1783
QY 1513 YLRCAALFFHYLLGVTPPEELHTNSAEGEYSALCSYLSLPTNLFLL---FOEYWDTVRPL 1569
DB 1784 -----QENSRAQENAKLCE-----TNIKELEROQY-----REQ 1813
QY 1570 LQRRCADPA--LLNCLKOKNTVVRPRKNSLIELPDDYSCLLNQASHFCRPSADDERK 1627
DB 1814 MQQGHMEANHYQKQCKLEDELIQKREVENLKKQKMDQOI-----KEHE 1857
```

QY 1628 HPVLCLFCGAILCSQNICQEIIVNGEEVGACIF-----HA--LHCKARGCAYP 1673  
Db 1858 HOLVLLQC-----EIOKSTAKDCTFKPDMFTVMEKCOHSGSLSSRNTGHLHP 1905  
QY 1674 APYLDEYGETDPLGRGNPLHLRSRERYKRLHLVWQOHCIIETAR 1718  
Db 1906 TP-----RSPLLRWTPQPLEEKW-QHRVVEQIPK 1935

## RESULT 13

US-09-727-384-6  
; Sequence 6, Application US/09727384  
; Patent No. US20020098511A1  
; GENERAL INFORMATION:  
; APPLICANT: Myriad Genetics, Inc.  
; APPLICANT: Heichman, Karen  
; APPLICANT: Cimbara, Daniel M.  
; APPLICANT: Bush, Angie  
; APPLICANT: Mauck, Kimberly  
; APPLICANT: Bartel, Paul L.  
; TITLE OF INVENTION: Protein-Protein Interactions  
; FILE REFERENCE: 2318-271  
; CURRENT APPLICATION NUMBER: US/09727,384  
; CURRENT FILING DATE: 2000-12-01  
; PRIOR APPLICATION NUMBER: US 60/168,377  
; PRIOR FILING DATE: 1999-12-02  
; PRIOR APPLICATION NUMBER: US 60/168,379  
; PRIOR FILING DATE: 1999-12-02  
; PRIOR APPLICATION NUMBER: US 60/185,056  
; PRIOR FILING DATE: 2000-02-25  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 6  
; LENGTH: 2139  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-727-384-6

Query Match 1.8%; Score 164.5; DB 9; Length 2139;  
Best Local Similarity 16.7%; Pred. No. 0.00027;  
Matches 241; Conservative 227; Mismatches 510; Indels 461; Gaps 54;

QY 174 GRAGTIKE--NSRPLNEEVIVQARKIFPSVIVKIVVEMTWEEKELPPELOIREKNR 230  
Db 697 GOAAVLKKAHHEATCRHEE-----KKQLQVKLEEKETHLOEKLRLOHEM- 732  
QY 231 YVCVLFNDEHSHYDVIVYSLQALDCELAEOHHTTAIDKEGRRAVKAGAY-----AAC 284  
Db 733 -----LAKRLTQAO--ASPERE-REGLOSSAWTEEKVRGLT 765  
QY 285 QEAKEDIKSHEN-----VSQHPLHVEVLHSEIMAHQKFAIRLGSWNMKIMSYSSDFRQIF 340  
Db 766 QELEG---FHQEQLTSLVEKHTLEELRKELEKHQRELOEG----- 805  
QY 341 COACLRPDSNPCLISRLMWDAKLYKGARKILHELIFSFEMEMEKYKLFAMEFVKY 400  
Db 806 -----REKMETECNRNRSQI---EAFQSDCKQVTERCESALQSLGEGRYR-----OE 849  
QY 401 YKOLQKEYISDDHRSISITALSVMFTVPTLARHLIEQNVISVITETLLLEVLPE-YLD 459  
Db 850 LKDQEQOREKSWQEFKEDELTOECAAQELKLTLEKREKTSLSVLTQEREMLEKTYKE 909  
QY 460 RNNKF--NFQGYSDQKLRGVYAVICDLKYLISKPTIWTPE---LRMQPLEGFRGFLKIL 514  
Db 910 HNSMVVERQQLQD-----LEDLRNVSTQOSLSLDQILELKSSHKRLREREVL 961  
QY 515 TCMQGMEEIRRVQGOHIEVDPDWEAAIAIQMOLKNILLMFOEWACDEELL--LVAYKEC 572  
Db 962 COAGASEQLASQRLERLEMEHDQER-----QEMSKLLAMENI 999  
QY 573 HKAVMRCSTSISSSKTVVQSGGHSLETKSYRVSEDLVSIHPLSRTLAGLHVRLSRL-G 631  
Db 1000 HKAT--CETA-----DRERAEMSTEISRLQS 1023

## RESULT 14

US-10-023-219-4  
; Sequence 4, Application US/10023219  
; Publication No. US20030032592A1  
; GENERAL INFORMATION:  
; APPLICANT: Myriad Genetics, Inc.  
; APPLICANT: Cimbara, Daniel M.

QY 632 AVSRLHEFVQFQEVVLPRLCLVLAQVVAEMWRRNGLSLSIQVFFYQDVKCRRE 691  
Db 1024 KIKEMOQATS-----PLSMLQSGCOVIGE-----BEV 1050  
QY 692 MYDKDIIMLOGLASLMDPNKFLLVATOR-YELAEAFNKTISTKODDLIKOYNLTLEMLQ 750  
Db 1051 EGDGALSLLQOGEOLLEENGVDLLSLQRAHQAVKVNKMATEISRLQORLOKL--EPGL 1108  
QY 751 VLIYIVGERYPVGVGNVTKBEVTMREIHLICIEPMPHSAIAKLNPENENNE---TGLEN 807  
Db 1109 VMSSCLDEPATEFEPGNTAEQTEQFLQNRRTKQVSGVTRRHVLSLDEDEVDRLSGTGS 1168  
QY 808 VINKVATFKPGVSGHGVYELKDESLDKDFNMFYHYSKTQSKAHMKOKKR-----858  
Db 1169 VQREQVKIESEASVEGSELEN-----SEETRTESWELKNQISOLQEQ 1213  
QY 859 -----RKQENKDEAL-----PPPPPEPCPAFSKVINLLNC-DIMMY 894  
Db 1214 MMLCADCDSASEKKQDILLFDVSVLKKLKLMLERPEASPKYLLIYEDVSRENDCLQEELR 1273  
QY 895 ILRTVFERAIDTDSNLWTEGMLQMAFHILALGLLEEKQOLQKAPEEEEVTFDFYHKASRLG 954  
Db 1274 MMETRYDEALENNKELTAE-----VFLQDELKKMEEVTFFLSLEKSYDEVKIE 1323  
QY 955 SSAMNIOML-----LEKLGIPOLEGQKDMITW-----ILQMPDT-----989  
Db 1324 NEGLNVLVLRLOQKIEKIQ--ESVVRQDCCLWEASLENIETPDGNILQNLQTLQECVP 1381  
QY 990 -----VKRLREKSLIVATTSGSESINKDEI-----TH-----1017  
Db 1382 RVRSVHHVIECKOENQYLEGNTQLLEKVKHAHEIAWLHGTIQTQHERPRVQNVILENT 1441  
QY 1018 -----DK-----EKAERKRAEAARLHROK-----IMAO 1042  
Db 1442 TLGFDQKHFOHOATTAELKTKLQELTKLKERVTILVQKQDVLSHGKEKEELKAMM 1501  
QY 1043 SALQKNFIETHK----LMYDNTSEMPGKEDSIMBEEESTPAYSD--YSRIALGKRGPSPVT 1096  
Db 1502 HDLIQITCSEMOQKVELLRYE--SEKLOQENSILRNEITTLNEEDSISNLKGLTGLNG---S 1556  
QY 1097 EKEVLTCILCOEQEVEKIENNAMVLSACVOKSTALTQHRGKPIELSGEALDPLFMDPOLA 1156  
Db 1557 QEE-----MMOKTETVKQEN-----AAVQKMVENLKKQISELKIKNQQLD-----1596  
QY 1157 YGTYTCSGCHVMHVAWCQKYFEAVQLSSQORIHVDLFDLESGEVLC-----1202  
Db 1597 -----LENTELQKNSQNOEKLOELNORLTMLCOKEKEPEGNSALEER 1639  
QY 1203 -----PLCKSLCNTVIPI-----IPLQPKQKINSEN---ADALAOLITLARW- 1240  
Db 1640 EQEKNLKELEKCKVQSSSTLVSSLEAESEVKIQTHTVQOENHLLKLEKMKQLHRCP 1699  
QY 1241 -----IQTVLARIAGYNIRHAKGENPIPIFN--QGMGDSTLEFHSILSFGVE--SSIKY 1291  
Db 1700 DLSDFQOKTSVSVSYNEKLLKEALSEELNSCVDKLAKSLLEHRIATMKQEQKSWBHQ 1759  
QY 1292 SNSIKEMVIL-----FATTIVRIGLVKPPDERPRVPLMTWTSCAFTIOA---IENL 1340  
Db 1760 SASLKSQVASQEKVONLEDTVQNVNLOMSRMSDLRVTQOEKALQEVMSLHKQLQNA 1819  
QY 1341 LGDEGEPFLGALONRQHNGLKALMQFAVORITCPVLIQKHLVRLLSVLPNIKSEDT 1399  
Db 1820 GGSWAPEIATHPSGLHNQOKRLSWDKLDHLMNEEQQLMQENERLQTMV-QNTKAE 1877



```

: APPLICANT: Heichman, Karen
: APPLICANT: Bartel, Paul L.
: TITLE OF INVENTION: Protein-Protein Interactions
: FILE REFERENCE: 2318-282-II
: CURRENT APPLICATION NUMBER: US/10/023,219
: CURRENT FILING DATE: 2001-12-20
: PRIOR APPLICATION NUMBER: US 60/256,983
: PRIOR FILING DATE: 2000-12-21
: NUMBER OF SEQ ID NOS: 4
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 4
: LENGTH: 2139
: TYPE: PRT
: ORGANISM: Homo sapiens
US-10-023-219-4

Query Match      1.88; Score 164.5; DB 15; Length 2139;
Best Local Similarity 16.7%; Pred. No. 0.00027;
Matches 241; Conservative 227; Mismatches 510; Indels 461; Gaps 54;

Qy 174 GRAGTIKE--NSRCPLNEEVIVQARKIFFSVTKYVVVEMTWEEKEKLPPELOLRKNNR 230
Db 687 GQAALVKEAHHEATCRHEE-----KKQLQVKEEBETHQLQKRLQHME- 732
Qy 231 YYCVLFNDEHSHSYDHVYISLQRALDCELAQAHTTTAIDKEGRRAVRKAGY-----AAC 284
Db 733 -----LKAULTQAAQ---ASFERE-REGLOSSANTEEKVVRGLT 765
Qy 285 QEAKEDIKSHEN----VSOHPLHVEVLHSEIWAHQKFAURLGSMWNKIMSYSSDFRQIF 340
Db 766 QELEQ----FHQELTSLVEKHTELEKELRKELLEKHQRELQEG----- 805
Qy 341 CQACLREPDSENCLLSRLMLNDAKLYGKARKTLHLELIFSSPFMEYKKLFAMEFVKY 400
Db 806 -----REKMETECNRRTSQI---EAQFSQCVKTERCESALQSGLEGYR-----QE 849
Qy 401 YKQLQKYSIDSDHRSISATLSVQMFTVPTLARHLIEEQNVISVITETLLEVLPE-YLD 459
Db 850 LKDLQEQOREEKQWFEKDELTEQCAEAQELLKETLKRKTTSLVLTQEREMLEKTYKE 909
Qy 460 RNNKF--NFGQYSODKLG RVYVJCDLKYILISKPTTWTER---LRMQFLEGFRSFKLIL 514
Db 910 HLNSMWYVERQOQLQD-----LEDURNVSETQOSSLSDQILLEKSKHRELREEREVL 961
Qy 515 TCMQGMEEIRQVQCHIEVDPDFWEAATAIQMLKNILLMPQEWCAQDEELL--LVAYKEC 572
Db 962 CQAGASQLASQRLERLEMEHQDER-----OEMMSKLAMENI 999
Qy 573 HKAVMRCSTSPISSSKTWVQSGHSLKTSYRVSEDLVSIHPLSLRSLAGLHVLRSL-G 631
Db 1000 HKAT--CETA-----DRRAEMSTEISRLQS 1023
Qy 632 AVSRLHEFVSPEDFQVEVLVEYPLRCLVLVAQVVAEMRRNGLSLISQVFFYQDVKCREE 691
Db 1024 KIKEMQATs-----PLSMLQSGCQVIGE-----EEV 1050
Qy 692 MYDKDITMLQIGASIMDPNPKFLLLVLQOR-YELABAFNKTISTKQDQLIKQVNTLIEMLQ 750
Db 1051 EGDGALSILQOGEOLLEENGDLVLSLQRAHEQAQVKNVKNMATSISLQQRQLKL--EPGL 1108
Qy 751 VLIYIVGERYVPGVGNVTKEVTWRETIHLILCIEPMPHSATAKNLPENENNE---TCLEN 807
Db 1109 VMSSCLDEPATEFFGNTAEQTEQPLQNRKTKQVEGTVRRHVLSOLEDDVDRLGSGTSS 1168
Qy 808 VINKVAITFKPGVSGHGYVELKDESLKDFNNFYHYSHYKSHQSAEHMQKR----- 858
Db 1169 VQREQVKIEESEA SVEGFSLEN-----SEETRTSESWEKKNQISQLQEQ 1213
Qy 859 -----RKQENKDEAL-----PPPPPPPCPAFSKVINLLNC-DIMMY 894
Db 1214 MMLCADCDDRASERKQDILLFDVSVLKKKLKM LERIPASPKYKLLIYEDVRENDLQBELR 1273
Qy 895 ILRVVFERRAIDTSDNLNLTGEMLOWAFHILALGLLEEKOOLQKAPBEEVTFDYHKASRLG 954

```

```

? FILING DATE: 23-OCT-1997
? APPLICATION NUMBER: US 08/329,154
? FILING DATE: 24-OCT-1994
? APPLICATION NUMBER: US 08/161,907
? FILING DATE: 02-DEC-1993
? APPLICATION NUMBER: US 07/985,321
? FILING DATE: 04-DEC-1992
? APPLICATION NUMBER: US 07/429,791
? FILING DATE: 31-OCT-1989
? ATTORNEY/AGENT INFORMATION:
? NAME: Ingollia, Diane E.
? REGISTRATION NUMBER: 40,027
? REFERENCE/DOCKET NUMBER: OPHD-01121
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (415) 705-8410
? TELEFAX: (415) 397-8338
? INFORMATION FOR SEQ ID NO: 6:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 2710 amino acids
? TYPE: amino acid
? TOPOLOGY: linear
? MOLECULE TYPE: protein
? SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-10-011-366-6

```

Query Match.	1.7%;	Score 156.5;	DB 15;	Length 2710;
Best Local Similarity	18.6%;	Pred. No. 0.0021;		
Matches	189;	Conservative 154;	Mismatches 354;	Indels 317; Gaps 44;

  

QY	288	KEDIKSHSENVSOHPLHVEV--LHSEIMAHQFALRLGSMWNKIMSYSDFRQIFCQAC	344
DB	355	KSEIFSKLENLNVSDDEIKIAFALGVI--NQALISKQSGYLNVLNVEQVKNRYQFNLQH	412
QY	345	LREPEQSENCLISRLMLWDAKLYKGARKILHELIFSFEMEMEYKKLFAMEFVKYKQL	404
DB	413	LNPALIESDNN-----FTDPTKIFHDSLFNSATAE---NSMFLTKAPY---L	453
QY	405	QKEYISDDHRSISITA-----LSVMFTV-PTL-ARHLIE---EQNVISVITE	448
DB	454	QVGFMPPEARS-TISLSPGAYASAYDFINLQENTIEKTLKASDLIEKFPENNLSQLTE	512
QY	449	TLEVLPEYLDNRNKNFQGYSDOKLGRVYA-----VICDLKYLILSK-PTIWT	497
DB	513	QETNSLWSFQASAKYQFERYKRDYTGGSUSEDGVDFNKNTALDKNYLLNNKIPSNVE	572
QY	498	RLRMQFLEGRSPRLKILTCMQGMEETRRQGVQHEIYDPDWEAAIAIQMLKNILLMFOEW	557
DB	573	E-----AGSKNVVHYIIQOGDDISYEATCNLFKNP--KNSIIQORNWNEGSAKSY--F	622
QY	558	CACDEELLVAYKECHKAVMCRSTSFISSSKTVVQSGH-----SLETKSYR	604
DB	623	LSDDGESIL---ELNK--YRIPERLKNKEKVYTFIGHGKDEFNTSEFARLSVDLSLNE	676
QY	605	VSEDLSVIHPLSRTAGLHVLRSLRGAVSRHLHEFVSFEDQFQVELVVEYPLRCLVLVAQV	664
DB	677	ISSFLDTIKLIDSPK-----NVEYNLLGC-----NMFESY-DFNVE--ETYPGKLLLSIMDK	724
QY	665	VA-----EMWRNRCLSLISQVFFYQDVKCREMY---	693
DB	725	ITSTLPDVNKNSTIGANQVEYVINSEGRKELAHGSKWINKKEAINDSLDSKEYIFPDS	784
QY	694	-----DKDIIMIQIGASLMDPNKPLLLVQ-----RYELAE	724
DB	785	IDNKLKAKSNIPCLASISEDIKTILLDASVSDPTKFIINLNKLNKINTESSICDVIYYEKL	844
QY	725	AFNKTISTKDDQLIKOYNTLIEMLQVLIYIVGERVYPGVGNVTKVEVMTREIHLLCIE	784
DB	845	PVKNIITHNSIDDLIDEFNLL-----ENVSDELYELKKL-----	877
QY	785	PMPHSAIAKNLPE-----NENNETGLENVINKVATEKFGVSGHGVYELKDESLKD	835
DB	878	-----NNLDKYLISFEDISKNSTSVREINK-----SNGESYIVVTEKRI--	919

Qy	836	FNWYFYHYSK-----TQHSKAEHMQKRRKQENKDEALPPPPPEFCPAFSKV	883
		:	
Db	920	FSKYSEHTTKEISTIKNSIITDVNGNLLDNITQDHTSQ-----	957
		:  :  :	
Qy	884	INLNCIDIMMYLRTVTFERAIDTD--SNLWTEGMLQMAFHILALGLEEKQLOKAPEEE	941
		:  :  :  :  :     :  :     :     :     :	
Db	958	VNTLNA---AFFIQSLDYDSNKNVDLSDTSVKVQLYAQLFSTGLNTIYDSITQV---	1010
		:  :  :     :     :     :     :     :     :	
Qy	942	VTDFYHKASRLGSSANN--IQMLEKXIKGIPQLEGOKDMTW--TLQMFDTVKRLREK	996
		:  :     :     :     :     :     :     :     :	
Db	1011	-----NLISNAVNDTINVLPITTEGPIPVSTILDGINLGAAIKELLDHDPULKK	1060
		:  :     :     :     :     :     :     :     :	
Qy	997	-----SCLVATTSSSESINKDEITHDKEAERKRAEAARLHROKIMAQMS	1043
Db	1061	ELEAKVGVLAINWSLSTAATVASIVGI-GAEVT-----IFLPIAGISACIP	1106
Qy	1044	ALQKNFIETHKLMYDNTSEMPGCKEDSIMEBESTPAVSDYSRIALGPKRGPSVTEKE-VLT	1102
Db	1107	SLVNN-----ELIHDKATSVVNYFNHLSSESKYGPILKTEDDKILV	1147
Qy	1103	CILQCEQEYKVIENNAMVLSACVOKSTALQHRGKPIELSGEALDPLFMDPDIA	1156
Db	1148	PIDDLVISEIDFNNSNKLGTG----NILAMEGSGHTVTGN-IDHFSSPSPSTS	1196

Search completed: September 25, 2003, 14:52:47  
Job time : 66 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 25, 2003, 14:51:16 : Search time 59 Seconds  
(without alignments)  
2826.381 Million cell updates/sec

Title: US-09-724-126A-19  
Perfect score: 9141  
Sequence: 1 AMEGNMADEEAGGTERMEIS.....EIARSQETNQMLFGFNWQLL 1734

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_76:\*

1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8480	92.8	1757	2 T14318	ubiquitin-protein
2	1886.5	20.6	1927	2 T25604	hypothetical prote
3	990	10.8	1958	2 T39808	hypothetical prote
4	866	9.5	2052	2 T37711	probable n-end-rec
5	691.5	7.6	1950	2 S12332	ubiquitin-protein
6	628.5	6.9	1941	2 T30554	ubiquitin-protein
7	606.5	6.6	1225	2 T48251	ubiquitin-protein
8	535.5	5.9	1872	2 S64851	probable membrane
9	507.5	5.6	2168	2 D88131	protein F10G7.10 [
10	292	3.2	795	2 T48252	eceriferum3 (CER3)
11	218	2.4	271	2 T40238	hypothetical prote
12	205.5	2.2	3187	2 JC5837	364K Golgi complex
13	190.5	2.1	3259	1 A56539	giantin - human
14	186	2.0	1410	1 A57013	early endosome ant
15	182	2.0	2442	2 T08621	centrosome associa
16	180.5	2.0	1413	2 T26467	hypothetical prote
17	179.5	2.0	2346	2 T13829	tptr homolog - frui
18	176	1.9	3225	2 I52300	giantin - human
19	175.5	1.9	1427	2 S22695	restin - human
20	175.5	1.9	1526	2 T41522	myosin ii - fissio
21	175.5	1.9	1830	1 S19188	myosin-v - chicken
22	175.5	1.9	2663	1 S12861	centromere protein
23	174.5	1.9	1009	2 C89910	hypothetical prote
24	174.5	1.9	1392	2 A43336	microtubule-vesicl
25	172.5	1.9	1780	2 T17272	hypothetical prote
26	168	1.8	1837	2 T41023	probable nuclear p
27	166	1.8	1163	2 G97236	Atfase involved in
28	166	1.8	1676	2 E71410	probable centromer
29	165	1.8	2649	2 A40937	bullous pemphigoid

30	164.5	1.8	1181	2 T01799	hypothetical prote
31	164	1.8	2954	2 T14156	kinesin-related pr
32	163.5	1.8	1339	2 H81307	restriction modifi
33	162	1.8	1269	2 F84730	probable myosin he
34	162	1.8	2335	2 T40186	probable phosphati
35	157.5	1.7	1642	2 T08880	NMDA receptor-bind
36	156.5	1.7	1496	2 T05634	hypothetical prote
37	156.5	1.7	1790	2 S67593	transport protein
38	156.5	1.7	2710	2 A37052	toxin A - Clostrid
39	156.5	1.7	2829	2 A42771	reticulocyte-bind
40	156	1.7	1853	1 A46761	myosin heavy chain
41	155.5	1.7	1216	2 A28821	1-phosphatidylinos
42	155	1.7	1025	2 E86355	hypothetical prote
43	155	1.7	1173	1 A53430	1-phosphatidylinos
44	154.5	1.7	2269	2 T28677	rhoptyr protein -
45	154	1.7	2469	2 H36812	hypothetical prote

## ALIGNMENTS

## RESULT 1

T14318

ubiquitin-protein ligase E3-alpha - mouse

N;Alternate names: N-recognin E3-alpha

C;Species: Mus musculus (house mouse)

C;Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999

C;Accession: T14318

R;Kwon, Y.T.; Reiss, Y.; Fried, V.A.; Hershko, A.; Yoon, J.K.; Gonda, D.K.; Sangan, P.

Proc. Natl. Acad. Sci. U.S.A. 95, 7898-7903, 1998

A;Title: The mouse and human genes encoding the recognition component of the N-end ru

A;Reference number: Z17977; MUID:98318583; PMID:9653112

A;Accession: T14318

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: mRNA

A;Residues: 1-1757 &lt;KWO&gt;

A;Cross-references: EMBL:AF061555; NID:g3170886; PID:g3170887; PIDN:AAC40165.1

C;Genetics: Ub1

A;Gene: Ub1

A;Map position: 2

Query Match 92.8%; Score 8480; DB 2; Length 1757;  
Best Local Similarity 91.0%; Pred. No. 0;  
Matches 1598; Conservative 64; Mismatches 67; Indels 28; Gaps 4;

QY 6 MADEEAGGTERMEISAEIPOTPORLASWDDQVDFYTAFLHHLAQLVPEIYFAEMDPDL 65

Db 1 MADEEMDGAERMDVSPPEPLAPQRPASWDDQVDFYTAFLHHLAQLVPEIYFAEMDPDL 60

QY 66 KOESVQMSITPLEWYLFGEPPDICLEKLKHSQAFQICGRVFKSGETTYSCRDCAIDPT 125

Db 61 KOESVQMSITPLEWYLFGEPPDICLEKLKHSQAFQICGRVFKSGETTYSCRDCAIDPT 120

QY 126 CVLCMDFQDSVHKHNRHKMHTSTGGGFCDCGDTGTEAKTGPFCVNHPEGRAGTTKENSRC 185

Db 121 CVLCMDFQDSVHKHNRHKMHTSTGGGFCDCGDTGTEAKTGPFCVNHPEGRAGTTKESLHC 180

QY 186 PLNEEVIVQARKTFPSVIKYVVENTIWEEEKELPELQIREKNERYICVLENDEHHSYDH 245

Db 181 PLNEEVIVQARKTFPSVIKYVVENTIWEEEKELPELQIREKNERYICVLENDEHHSYDH 240

QY 246 VIYSLQALDCELAELAQHTTAIDKEGRRVAKAGAYAACQAKEDIKSHSNVSOHPLHV 305

Db 241 VIYSLQALDCELAELAQHTTAIDKEGRRVAKAGAYAACQAKEDIKSHSNVSOHPLHV 300

QY 306 EVLHSEIMAHQKFAIRLGWSNMKIMSYSSDFRIFQCOACLRPEEDSENPCILISRLMWD 365

Db 301 EVLHSEIMAHQKFAIRLGWSNMKIMSYSSDFRIFQCOACLRPEEDSENPCILISRLMWD 360

QY 366 KLYKGARKILHEIFSSFFMEMEYKFLFAMEFVYKQLOKEYISDDHRSISITALSQV 425

Db 361 KLYKGARKILHEIFSSFFMEMEYKFLFAMEFVYKQLOKEYISDDHRSISITALSQV 420

```
QY 426 MFTVPTLARHLIEQNVISVITETLLEVLPEYLDNRNKNFQGYSDQKLGRRYAVICDLK 485
Db 421 MFTVPTLARHLIEQNVISVITETLLEVLPEYLDNRNKNFQGYSDQKLGRRYAVICDLK 480
QY 486 YILISKPTIWTBLRQWQLEGRFSFKILTCMGMEEIRQVGOHIEVDPDWEAAIAIQM 545
Db 481 YILISKPTIWTBLRQWQLEGRFSFKILTCMGMEEIRQVGOHIEVDPDWEAAIAIQM 540
QY 546 QLNKILLMFQEWACDELLLVAYKECHKAVMRCSTSFISSTTVVQSGHSLTKSYRV 605
Db 541 QLNKILLMFQEWACDELLLVAYKECHKAVMRCSTNFMSSTTVVQLCGHSLTKSYKV 600
QY 606 SEDLSIHPLSTLTLAGLHVRLSRLGAVSRHLHFVDFEDQVEVLVEYPLRCLVLVAQVV 665
Db 601 SEDLSIHPLSTLTLAGLHVRLSRLGAVSRHLHFVDFEDQVEVLVEYPLRCLVLVAQVV 660
QY 666 AEMWRNGLSLISQVYQDVCKREEMDKDITMLQIGASIMDPNPKFLLVLQRYELAEA 725
Db 661 AEMWRNGLSLISQVYQDVCKREEMDKDITMLQIGASIMDPNPKFLLVLQRYELTDA 720
QY 726 FNKTISTKDQDLIKOYNLTIEEMQLVLIYVGERYPVGVNVTKEEYTMREIHLGLCIEP 785
Db 721 FNKTISTKDQDLIKOYNLTIEEMQLVLIYVGERYPVGVNVTKEEYTMREIHLGLCIEP 780
QY 786 MPHSATAKNLPENNETGLENVINKVATPKFGVSGHGYVELKDESLKDFNMFYFHYSK 845
Db 781 MPHSATARNLPENNETGLENVINKVATPKFGVSGHGYVELKDESLKDFNMFYFHYSK 840
QY 846 TQHSKAHMKKRRKQENKDEALPPPPPECFAPFASKVINLLACDIMVYTLRTVFERAID 905
Db 841 TQHSKAHMKKRRKQENKDEALPPPPPECFAPFASKVINLLACDIMVYTLRTVFERAID 900
QY 906 TDSNLTEGMLQMAFHIALGLLEEQQLQKAPEEVTFDYHKASRLGSSAM---NIQM 962
Db 901 TESNLTEGMLQMAFHIALGLLEEQQLQKAPEEVAFDYHKASRLGSSAMNAQNIQM 960
QY 963 LLEKLKIPOLEGOKDMITWILQFDTVTKRLREKSLIVATTSGSESIKNDIETHDKEA 1022
Db 961 LLEKLKIPOLEGOKDMITWILQFDTVTKRLREKSLIVATTSGSESIKNDIETHDKEA 1020
QY 1023 ERKKAFAARLHOKTMAQSALOKNFIEHKLMYDNTSEMPGKEDSIMBEESTPAYSDY 1082
Db 1021 ERKKAFAARLHOKTMAQSALOKNFIEHKLMYDNTSEMPGKEDSIMBEESTPAYSEA 1080
QY 1083 SRIALGPKRGPSYTEKEVLTICLQEBQEVKIENNAMVLSACVQKSTALTQHRGKPIELS 1142
Db 1081 SRIALGPKRGPAVTEKEVLTICLQEBQEVKLENNAMVLSACVQKSTALTQHRGKPDVHL 1140
QY 1143 GEALDPLFMDPLAYGTYTSGCHVHVAWCQKYFEAVQLSSQRIHVDFDLESGEYLC 1202
Db 1141 GETLDPLFMDPLAHGTYTSGCHVHVAWCQKYFEAVQLSSQRIHVDFDLESGEYLC 1200
QY 1203 PLCKSLCNTVPIPILOPKINSENADALAQLTLARWIOVLARISSYNIHRAKGENP- 1261
Db 1201 PLCKSLCNTVPIPILOPKINSENADALAQLTLARWIOVLARISSYNIHRAKGENA 1260
QY 1262 IPIFFNGMGDSLEPHSILSFGVSESIKSNKEMVILFATTIYRIGLKVPPDERDPR 1321
Db 1261 VPVLFNGMGDSIFEFHSILSFGVSSVKYSNISKEMVILFATTIYRIGLKVPPDELDP 1320
QY 1322 VPMLTWSTCAFTTQATENLGLDGKPLFGALQNRQHNGLKALMQFAVAQRITCQVLIQK 1381
Db 1321 VPMWTWSTCAFTTQATENLGLDGKPLFGALQNRQHNGSLKALMQFAVAQRATCQVLIHK 1380
QY 1382 HLYRLLSVLPNLSKEDTPLLSTDLPHVLGAVLAFPSLYWDDPVDLPQSSVSSYNHL 1441
Db 1381 HLYRLLSVLPNLSKEDTPLLSTDLPHVLGAVLAFPSLYWDDPVDLPQSSVSSYNHL 1440
QY 1442 YLFHLITMAHMLQILLTIVDR---GLPLAQVQDSEEAHSAFPAISQVTSIGCDI 1497
Db 1441 YLFHLITMAHMLQILLTIVDR---GLPLAQVQDSEEAHSAFPAISQVTSIGCDI 1490
QY 1498 PGWYLVWSLKNGITPYLRCAALFFHYLLGVTPPELHTNSAEGEYSALCSYLSLPTNLF 1557
```

## RESULT 2

T25604

hypothetical protein C32E8.11 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C:Accession: T25604

R:Gattung, S.

submitted to the EMBL Data Library, February 1997

A:Description: The sequence of C. elegans cosmid C32E8.

A:Reference number: Z20056

A:Accession: T25604

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1927 &lt;GAT&gt;

A:Cross-references: EMBL:U08308; PIDN:AB42328.1; GSPDB:GN00019; CESP:C32E8.11

A:Experimental source: strain Bristol N2; clone C32E8

C:Genetics:

A:Gene: CESP:C32E8.11

A:Map position: 1

A:Introns: 56/3; 157/3; 255/1; 469/1; 578/3; 832/3; 907/2; 1165/2; 1218/2; 1275/2; 14

Query Match

Best Local Similarity 20.6%; Score 1886.5; DB 2; Length 1927;

Matches 553; Conservative 305; Mismatches 656; Indels 513; Gaps 67;

QY 102 QLCGRVFKSGTETYSRCDADPTCVLCMDCFDQSVHKNHRYKMHMTSTGGFCDCGDTA 161

Db 14 QICGHVFKNGELTYTCLCATDGTCTVCLQCEVSIHSHKHYKMHSSSGYCDGDADA 73

QY 162 WKTGPFVYNHEPG-----RAGTIKENSRCPLNEEVIVQARKIPSPVIKYVEMTWEE 214

Db 74 WTEGYACANHEKDKDEEAVALPELKK--RCEQLVEILQ---FSLSMITHKDDLKLEI 128

QY 215 EKELPPELOIREKNERYCYVLFNDEHSHVDHYIYSIORALDCELAQLHTTAIDKEGR 274

Db 129 FEKKMK--VTNEAQOYLTVLYNDETHYVESVLEKVIHCTKQAMLVATIVDREGS 186

QY 275 AVKAGAYAAOCBAKEDIKSHS-----ENVSOH--PLHVEVLHSHIMAHQFALRGSW 325

Db 187 AVKLGSKADCTYKAKDDVQKRTARDPTSIRRSNNHLPLSVKVMDDTLFALQNFISILLTW 246

QY 326 MNKIMSYSDFRQIFCQACL-----NKTSMYSDFRQIFCQACL----- 345

Db 247 LNTQMDVPPLPREIVEIGEILLSSKFALKKNTYRKMSQDQRLVAGIIRNVWLPDDEEEL 306

QY 346 -----REEP-----DSENPCLI 357

Db 307 FALDGRMDVDEMDDDDIGEALQMEIDADEEETAAALAGVSEHQSGPGRSDSFTTFL 366

QY 358 SRLMLWAKLYKGARKILHELHLSFFMEMEKYKLFAMEFVYKYLOKQYISDDHRSI 417

Db 367 ENILLQDTQMKGAGRSIHLQMLMRTVMFYDQKVRPAKAFMLHYNEIEDFIKDDHMDV 426



Db 217 ESTLQDEKTSRLSENKYGIDDDSCNMYSLVLWDEKHSFKQFYEQIITALL--ELPNNVFG 274  
Qy 262 QLHTTADTKGRRRAKAGACAAQEKEDIK---SHSENVSQHPHVEVLHSEIMAHOKF 318  
Db 275 KMANIINDIGR-----ACTVETNIKELKIGOKLAQINLAVSIRMRDIFREES 325  
Qy 319 ALRGLSNMKNMYS-----SDRQIFCOACLR----- 346  
Db 326 CAVLEWADIAGSSICGKKNYFSSVICKELVRPWCNGLHNSDITFRLSLRSLALPEIVA 385  
Qy 347 -EEDP-----SENPCILISRLM----- 361  
Db 386 IDSPDIFLNEDHINSSGSDTSSTMLETDESSIHSHRWYPSNSLPDVLASYASRVDFYFF 445  
Qy 362 LMDAKYKGARKILHELIFSSFFMEMEYKKLFAMEFYKQYKQLOKEYSIDHDHSISITA 421  
Db 446 LYDLKLKWSLRYKLQELVGLVFTQPGFKEIMGARIAISYRRLAELFLLLDREPHSVIF 505  
Qy 422 LSVQMFVPTPLARHLIEQNVISVITETILEVL-----PEYLD-----RNKNFNQGY 469  
Db 506 FMSQIFTVAOVAKLLVTEYDFLTINATLYTFTYKKLNTPNYVDQHAMIITDSAAHF-- 563  
Qy 470 SQDKLGRVAVICDKVLILSKPTIMTERLR--MOFLEGRFSFKILTCMOGMEIIRVOV 527  
Db 564 -----SRYTHIFHHIQFMLSIPCV-AEIVREDLKLKQYADFFNL-----FGCMCPYTRAV 614  
Qy 528 GOHIEVDPD-WEAAIATOMQL-----KNILLMFQEW-----CACDEELLVAYRECHKAV 576  
Db 615 SQHVENWDSMWYLVNLSQVAKLCHVGNVEMELNKLANAINYLISLILYPKARNES 674  
Qy 577 MRCTSTISSKTVQSCGHSLETKSYRVEDLSIHLPLSRTLAGHLVRLSRLGAVSRL 636  
Db 675 W-TNTESLTGTITVDER-GNS-KLIEYDIALQPVSFHPLHLLVYL-----L 719  
Qy 637 HEFVSFEDF-----QVFLV--VEYPLCLVLVAOVAAEMRRNGLSLISQVFFYQDVKCR 689  
Db 720 SFYVERDNYKLLWTQDOLLAVTDPLRCVACLWSQRAKLWTRNGTTIURDQAHHRNLSFH 779  
Qy 690 EEMVDKDIIMLQIGASLMDPNKFLLLVLQRYELAEAF--NKTIKQDQDLKQYNTLIEE 747  
Db 780 EYTFDLDVLLQLTLTYGDPDAILPSFISRFQEDQMYGREFFVPHKYD-VSQVTIMTEE 838  
Qy 748 MLQVLIIVIGERYVPGVGNVTKEVMTREIHLICIEBPMPSHAIKPNENNETGLEN 807  
Db 839 FLILLISIVCNTAVLDHWDITR--IEYGAHILCFRPLPYSEIKTKRCEHLLHOKPES 896  
Qy 808 VINKVATFK-PGVSCHGVYELKDESKDFNMYHYHYSKTQHSKAEHMOKKR-RKQENKD 865  
Db 897 TLKVAFTRNAEINGSGSTFKDEYFDYVDPFNIHYSRNQREAEANILRRRYSKQSHK 956  
Qy 866 EALPPPPPPFCPAFSK--VINLLNCIDIMYIL-RTVFERAIDTDSNLWTEGMLQMAPHI 922  
Db 957 --LESVVEEYHPILHSHNITPILOSFSFGILMHTIYVAYIYDQKLEGLVNTALHA 1014  
Qy 923 LALGLEEKQOLQKAPAEVETFDYHKASRLGSSAMNIQMLEKLGIPQLEGQD----- 978  
Db 1015 CLLVLMSEK-----GSEPIFSKKICE--NRFPVVEGLQYCNYS 1050  
Qy 979 ----MITWILQM-----FDIVKRLREKSLIV-----ATTSGSSIKNDE 1014  
Db 1051 PDVTLFVLVQCMKNHRNFVYVK-----EKISLIMKILKSEVPILLYBPVVAETLSISSKTIQV 1107  
Qy 1015 ITHDKKAERK-RKAEARLHROKIMQMSALOKNFLETHKLMYDNTSEMPGKEDSIMEE 1073  
Db 1108 SLSDAEQOEQHLAKVRWAKERQARIMEQFRQONKFNLENHALFASDCM-----DEADEF 1163  
Qy 1074 ESTPAVSQDYRIALGPKRGPSVTEKEVLTCTLCQEEQEVKLENNAMVLSACVQKSTALT- 1132  
Db 1164 SVTSSVS--TKLFLDP-----PIDTCLLCOEELKDKRPYGLTFV---VLRSVLR 1209  
Qy 1133 -----QHRGKPIELSG-----EALDPL----- 1149  
Db 1210 FPADDANYVSEVLDPDLSLOHEIOERPFGLAGKRKKVLDSTPEAYDYDNYVYKKGKGNELHQ 1269

Qy 1150 -----FMDPDLAYTYTTCGCHGVHVAWCQYFEAVOLSSQ---QRTHVLDLDESSE 1199  
Db 1270 LKDSFNGFPDQDRGLHATGCGHPMHIDCFKNHIAVTATVTRANPRYRNP--HNLSMKE 1327  
Qy 1200 YLCPLCKSLCNVTIPIIPLQOKINSENADALAQLLTARWI-----QTVL-- 1245  
Db 1328 FLUCPLCKALCNTIFPILWRPKEEINFQEAQVLT--APLKNVLVSKTFSFNKDLNQQLLDI 1385  
Qy 1246 -----ARISGYNIRHAKGENPIPIFFNCGMGDSTLEFHSILSGVSESSIKYSIKEMV- 1299  
Db 1386 ETPSPSEHTQSYNL-----NLLDV-LQHTLRDSLKDIYT-LNTGADNS---SDNVEENAD 1434  
Qy 1300 ILFATTIYR-----IGLKVPDPDERDPRVMTWTSCAP-----TIQAIENLGGDE 1344  
Db 1435 NLFQSVLDHVFHFKSVVNEVPADER-----LAISDDIFELYRRLDDVIDLNSLSYSD 1488  
Qy 1345 GKPLFGALONROHNGLKALMQFAVAQ-----RITCPOVLI-----QKHLVRL 1387  
Db 1489 FIPVNGKL-----HNVVK-LFSYSLCQVEASTRGHIKCSSIPADIWVHNLGKNQOVFLRIL 1543  
Qy 1388 SVVLNPNIKSEDPCLLSIDLPHVLVGAFLAF-----PSLYWDDPV--- 1427  
Db 1544 S-----ESIKTYTLLC--AHDQKRIGSGIOEPFISFCQOKRIFGRLLPSL--DSPYTKSI 1596  
Qy 1428 -----DLQPSVVS-----SSYNHL-YLFHLITMAHMLQIILLTVDTLPL 1465  
Db 1597 TDDRVEPLLVKDTREFEAEASVGLLSCDESFHYLTQLYVTADIVRNLTLLSORSNL-L 1655  
Qy 1466 AQVQEDSEEA-----HSASSEFAISQYTSISICDI-----PG--WYLWVS 1505  
Db 1656 KOMESVEFEAFDYEQLKGEHLVIOIWKSLRVGAGLINFDCTDDEDLNPHLFLTYKL 1715  
Qy 1506 LKNGITPYLRCAALFFHYLLGYTPPEELHTNSAEGYSALCSYLSLPTNLF---LFOEY 1562  
Db 1716 LERFSLIFLRKCALLWYCRYGVSEFOTPNLNFQNSLSRLQTKMHIPGVIELSNHLCILTA 1775  
Qy 1563 WDTVRLQLORCADPALLNCLNOKNTVVRYPKRR-----NSLIELPDYSCLLNQASHFR 1618  
Db 1776 SSTEWSLIKHWK-----NFFTETGCLDFPRAYPGIVELVSLPYELDKVFELLARRC 1829  
Qy 1619 PRSADDERKHPVLFCGAILCSQNTCCOEIVNGEVEGACIFALHCKA-----RGC 1670  
Db 1830 SKCL-TEPMEPAICLFCGKLLCFQSHCCS--FNG--IGECNLHMQQCADIGIFLIVKCK 1884  
Qy 1671 AY-----PAPYLDEYGETDPLKRGPNPLHSRERY-RKLHLVWQOHCIEIAR 1718  
Db 1885 ALLYLNPVGVGSFSAFPLDAYGETDLGLRRGRSQYLSQKRYDETIVTMWMLNGSIPSYAR 1944  
Qy 1719 SOETNOMLFGFNQQLL 1734  
Db 1945 QLDANPDTCG--WETL 1958

## RESULT 4

T37711

probable n-end-recognizing protein - fission yeast (Schizosaccharomyces pombe)  
C:Species: Schizosaccharomyces pombe  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999  
C:Accession: T37711  
R:Murphy, L.; Harris, D.; Wood, V.; Barrell, B.G.; Rajandream, M.A.  
submitted to the EMBL Data Library, June 1997

A:Reference number: Z21738

A:Accession: T37711

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-2052 &lt;MUR&gt;

A:Cross-references: EMBL:Z97208; PIDN:CAB10108.1; GSPDB:GN00066; SPDB:SPAC15A10.11

A:Experimental source: strain 972h-; cosmid c15A10

C:Genetics:

A:Gene: SPDB:SPAC15A10.11

A:Map position: 1

Query Match				9.58; Score 866; DB 2; Length 2052;			
Best Local Similarity				20.28; Pred. No. 5.3e-45;			
Matches				443; Conservative 320; Mismatches 724; Indels 702; Gaps 90;			
QY	67	QESVQMSIPTLEW-YLFGEDP-----DICLEKLKHS-----GAFOLCGRVF	108				
DB	43	QEVFLSLLNEDNWKYFLKEKGAIITSDPRLSRLOHSEPECAQELQDKRSGSKVCGHVF	102				
QY	109	KSGETTYSRCDCAIDPTCVLCMCFQDSVHKHNRKMYKMTSTG-GGFCDCGDTEAWKTGPF	167				
DB	103	RAGEVIYRCKNCGLDNTCVLCAPCFATNHEGHEHETHVSIYSTSGICDCGDGPEAWNVDLN	162				
QY	168	CVNHEPCRAGTIKENSRCPLNEEVIQVARKIPFSVIKYVVMETIW-----	212				
DB	163	CKTH-----NVPDDEQKPEEVI-----PLEQHSRTTHILLDFILDFVFCSPV	209				
QY	213	-----EERKELPELQIREKNERY-----YCVLFNDEHHSHYDVIYS	249				
DB	210	NLKAQSTVGSILADEAS-----RLSSAKYGVADRPCNVFRVMLWDEVHFFDAVGS	262				
QY	250	LQALDCELAELQHTT-AIDKEGRRAVKAGA-----	280				
DB	263	VLEALDSNTAFGLEVAQVRDSTIGRAVATSSVHEAIRIANAIKENLANVNRVARTDFF	322				
QY	281	-----YAA-----COE-----	286				
DB	323	REDICGILWFDLLESHVCYFADYLQIIVCDEILKNWSPGLEKPAKPEVNFNNLPLEI	382				
QY	287	-----AKEDIKSHENVS-----QHPL-----	303				
DB	383	VNDDSEDDIYAAEELLDVIANLODETVRIANLGDEDEADMTDPTTAGDHPDLD	442				
QY	304	-----HVEVLHSEIMAHQF-----ALRGSWMNKMISY-----	333				
DB	443	NDVNDLDFETEREDIDLTDEVMEETENEAADYPGVNRNTRQDDVQDISMETESQNE	502				
QY	334	-----SDFRQIF-----	340				
DB	503	TDESQNTENVYNPQTHTPVPIPTATQDVVTIRPFNSQLNLRNLRQIINARRPRPAV	562				
QY	341	COACLREE-----PDS-----ENPCLISRL--MLWDAKLYKGARKTLHELIFSSFF	384				
DB	563	CQVSLREDYKSPHPPPSSYSVESPSSILRLDYFLFLDLKFWKRLRGLLSKLYVVPFN	622				
QY	385	MEMEYKKLFAMEVYKYLQKEYISDDHRSISITALSQVMEFTPTLARHLIEEQNVIS	444				
DB	623	RNLLFRLMGIRVHYRSLATAFLADREPDSHVMFLSVQFTTSLAEAVKDYDELT	682				
QY	445	VITETLLEVL-----PEYLDNRNKNFNQGYSD-----KLGRVYAVICDLK	485				
DB	683	NLNATTLSSLTQSNRPSTL-----FSSDIEVTPTIQLNRQVLKTRTYNLFSDLG	732				
QY	486	YIL-----ISKPTTWELRMOFLGPRSLKILTCMQGMEIEIRVOGQHIEVDPDWEAAI	541				
DB	733	YLQHPQVKLVVDDTRYHQYIDLRFV-----QGVIPQORAILSHVQ-----WDFPH	781				
QY	542	AIQOMQLNILLMFOEWCACDEEL-----LLVAYKECHKAVMRCSTSFSSSKTVVQSCGHS	597				
DB	782	G-----KNILFVMQVRAMLNNTVSSCFTQAPYERLFFYAI-KCIITSIHPKLDI---AES	832				
QY	598	LETKS-----YRVSEDLVSIHPLSRLTAGLHVRLSRGLGAVSRHLHFVS	641				
DB	833	LEPLSCIPSSSLTNFTQPLVPFVSRRDPISFYHP-----LHWMLSNLFSYCRVDASSH	885				
QY	642	FEDFQVEVLVEYPLRCLVLVAQVVAEMRRNGLSISQVFPYQDVVKCREEMYDKDIIMLQ	701				
DB	886	WDKDTLLALDHLRVCVLLAQDCNLWIRNGRSILLTDAFYROLNNIEVSYDKDILAIQ	945				
QY	702	IGASLMDPNKFLLLVQRYELAE-APNKTIISTQDQDLIKQYNTLIEBMLQVLIYIGERY	760				
DB	946	TILMFVDPNVLNAVYQREFTDWLYNATYNEHPNDYTERIPAMLCMKMLELLALITER-	1004				
QY	761	VPGVGNVTKEVMTREIHLCLTEPMPHSAIAKNLPENENNETGLENVINKVATFKKP-G	819				
DB	1005	-EQLIHVDIIDIIRTLAQQLCFGLPLAYSALLSTISSNLVESLSPDKIREETSYKAPDG	1063				
QY	820	VSGHGVTELKDESLKDFNMYYHYHYSKTOHSAEHMOKRRKQEN--KDBALPPPPPPPPFC	877				
DB	1064	LHDFGVYSLKDEYDLDVDPYFHYHKNREESDTILKRLAKKNVSAESIIEPKIRFL	1123				
QY	878	-----PAFSKVINLLNCIDIMMYILTRVFERAIDTDSNLWTGMLQW-AFHILALG--LLE	929				
DB	1124	EKQGHDIFFAANASTFSLIIF-----RAIE-----YALVOAESGSSDIGNTIILG	1169				
QY	930	EKOOL-----QKAPBEEVTFDFYHKA-----SRLGSSAMNIQMLLEKLKIPOL	973				
DB	1170	DALQCLLSMKIHEFSKSNDFCSRCAERYPTDSSIMREFGSAFYLAELCFAILKSPKY	1229				
QY	974	EGQKDMITWI---LQMF--TVKRLREKSLIVATTSSGSEIKNDEITHDKAEKRKA	1028				
DB	1230	KDVHVKNVAVLAGLQKNDPSAYSNNLEATHFELSTSS-----TSDSNEIEKTQEK	1280				
QY	1029	EAARLHRO-KTMAQMSALOKNFIETHKLMYONTSEMPGK---EDSIMEEESTPAVSDYSR	1084				
DB	1281	KRLALEKOKKIMQOFRDQOASFLA-----QNTDFDIDQDQTEDEVTTTEPEEVEKYHEH	1334				
QY	1085	IALGPKRGPSVTEKEVLTCILCOEVEVK-----IENNAMVLSACVQKSTALTQHRG	1136				
DB	1335	I-----RG-----NCLLCQECNDQAPYGVIGIIGSSLLRKTDVHSEIILDEIYS	1380				
QY	1137	KPIELS-----GEALDPLFMDPD-----LAY-----GTYTSGCHVMHVAWCQY	1176				
DB	1381	VPPNLDRESHSRPFQKDYTVVFNRSKORLLSAPPGNIRGVFVSGCGLHMLGCFKNY	1440				
QY	1177	FEAVOLSSQQRHVDLFDLESG--EY-----LCPCLSKLONTVPIIPILOPOKTN	1224				
DB	1441	YVARSMYRN-----DVTAGLSYKYYKSTAKFFMCPCLRSLSNLLP-MPQIPKM--	1489				
QY	1225	SENADALQALLTLARWIQTVLARISYNIIRHAKGENPIPIFNQGMGDSITLFSHSI---	1280				
DB	1490	CLNIDTLNPRSMNGWLEEI-GTWSSSSFEYQ-----LVRSLSUDTKDTRSCFLRP	1540				
QY	1281	-----LSFGVSSSIKYSN--IKEMVILFATTIYRIGLKVPPDERDPRV-PM	1324				
DB	1541	WINSKIISAMLARUKIADGALIDQNNRDVSDLYDRYCETT-KLANKLVKGSTFNVSPH	1599				
QY	1325	LTWSTCAFTIOAIENLLGDEGP-LFGALQNRHNGKAL-MQFAVAQRITCPOVLIQKH	1382				
DB	1600	DLNLSLAYTVSSLE--VSQRCSKPGQSGATRSWFNELGFLTSLFPLTSDT-----	1648				
QY	1383	LVRLSVVLPNIKSEDTPCLLSIDLHVLVGAVLAFPSLY-----WDDPVDLOPSSV	1434				
DB	1649	---VLKVCYDQIIRKSDQOALLMESQKLLVCKIFYRHSQKMLNRMGRMSDHDQOPFL	1705				
QY	1435	SSSY-----NHLYLFHLITMAHMLQILLTVDTGLPLAQOED---SEE	1474				
DB	1706	SNTDFDFVKISSMLIFGQDNILYYVKLFYLSSECKTIISMIKVVADSSVVPDLTINTS	1765				
QY	1475	AHSASSFAEISQVTSIGSGIDPCWYLMWSLKNGI-----TPYL	1514				
DB	1766	QOSKQFY-----ILC--KNVLLWSSNNIEILDDSNLLRLMSLYEKSLSPL	1813				
QY	1515	RCAALFFHYLGVTPP--EELHTNSAEGEYSALCSYLSLPTNLLFLQFYW-----DT	1565				
DB	1814	RRVALVYCMFDISLEFNEFSNNEDESELERLSKLKVPV-----LQELYSQMSSDENQ	1868				
QY	1566	VRPLLQRCADPALLNCLKQKNT-----VVRYPRKNSLIE-LPDDYSCL	1609				
DB	1869	ILELIAGMCEHLA-----QNTWGDSTISLEYPGIYELVLKPLHLELIDMSQMSVCCM	1921				
QY	1610	LNQASHFRCPRSADDERKHPVLCFLCGAILC---SONICQEIIVGEEVGACIFHALHC-	1665				
DB	1922	CHKTPIL-----PAICMLCGSVICFNARQNTVSSRRLTGE-----CNKHAATCT	1965				
QY	1666	-----KA-----RGCAYPAPYLDEYGETDPLGKRNGLHLSRERYR-KLH	1704				



Db 1966 GSVGIFITKACGILLLDISINTGTIMPTPYLDIHGETDLQLRGCGPQFLNQKRYDFVVR 2025  
Qy 1705 LVWOQHCHIEIARIQSQTQNMQLFGFNWOL 1733  
Db 2026 EQWLROTVALQKMARHMDMTQM-----NWRM 2051

## RESULT 5

ubiquitin-protein ligase (EC 6.3.2.19) - yeast (Saccharomyces cerevisiae) (strain S288C)  
S12332

A:Alternate names: protein G7168; protein YGR184C

C:Species: Saccharomyces cerevisiae

C:Date: 16-Sep-1992 #sequence\_revision 16-Sep-1992 #text\_change 03-Jun-2002

C:Accession: S12332; S64502; S64498

R:Bartel, B.; Wuening, I.; Varshavsky, A.

EMBO J. 9, 3179-3189, 1990

A:Title: The recognition component of the N-end rule pathway.

A:Reference number: S12332; MUID:91006011; PMID:2209542

A:Accession: S12332

A:Molecule type: DNA

A:Residues: 1-1950 <BAR>

A:Cross-references: EMBL:X53747; NID:g4743; PIDN:CAA37779.1; PID:g4744

A:Experimental source: strain S288C

R:Arroyo, J.; Garcia-Gonzalez, M.; Garcia-Saez, M.I.; Sanchez-Perez, M.; Nombela, C.

submitted to the Protein Sequence Database, May 1996

A:Reference number: S64499

A:Accession: S64502

A:Molecule type: DNA

A:Residues: 1-1950 <ARR>

A:Cross-references: EMBL:Z72969; NID:g1323325; PIDN:CAA97210.1; PID:g1323326; MIPS:YGR184C

A:Experimental source: strain S288C

R:Hebling, U.; Hofmann, B.; Delius, H.

submitted to the Protein Sequence Database, May 1996

A:Reference number: S64003

A:Accession: S64498

A:Molecule type: DNA

A:Residues: 1615-1950 <HEB>

A:Cross-references: EMBL:Z72969; MIPS:YGR184C

A:Experimental source: strain S288C

C:Genetics:

A:Gene: SGD:UBR1; PTR1

A:Cross-references: SGD:S0003416; MIPS:YGR184C

A:Map position: 7R

C:Superfamily: ubiquitin-protein ligase

C:Keywords: ligase; protein degradation

Query Match 7.6%; Score 691.5; DB 2; Length 1950;

Best Local Similarity 20.0%; Pred. No. 3.3e-34;

Matches 409; Conservative 321; Mismatches 734; Indels 581; Gaps 86;

Qy 12 GGTERTMSBELPQTTPORLASWMDQVDFYAF-LHHLAQLVPEYFAE-----MDPDL- 64

Db 33 GPTERADMSRALKEFIYR-----YLYFVINSNGENLPTLFNAHPKQLSNPELT 81

Qy 65 ---EKQESVQMSIFTPLEWYLFGE-DPDICLELKHGAFQLCGRVFKSGTTCYSCDC 120

Db 82 VFPDSLEDAVIDKITSQOTIPFYKIDSRIGDVHKHTG--RNCGRKEKIGEPYRCHEC 139

Qy 121 AIDPTCLVCMDFQDSVHKHURY--KMTSTGGGFCDCGDTAEWKTGPFQ-----VN 170

Db 140 GCDTCLVLCIRFPKDHVNHVCTDICTETSTGICDCGDEAMNSPLHCKAEQENDIS 199

Qy 171 HEPGRAGTIKNSRCPNEEV-IVQARKIFSVIKYVEMTWEEKELELPPELO----- 223

Db 200 EDPATNADIKED--VWNSDVNIALVELVLAEVDFYFD--VFNONIEPLPTIQDKITK 255

Qy 224 IREKN-----ERYQVL 235

Db 256 LREMTQOQKMYERAQFLNDLKYENDYMDGTTTAKTSPNSPEASPLAKIDPENYVII 315

Qy 236 FNDEHSHVDHYVLSQALDCELAQHLTAIDKEGRRAVKA-----GAYAACOE- 286

Db 316 YNDEYHNYSQATTALRQGV-DNVHIDLTLTRIDGEGRAMLKCSQDLSSVLGGFFAVQTN 374

Qy 287 -AKEDIKSHSENVSOHPLHVEVL---HSEIMAHQKFAALRGLSGMMNKIMYSDDF----- 336  
Db 375 GLSATLTSWSEYLHOETCKYIILWITHCLNIPNSFQTFRNMGMKTL--CSEYLNATEC 432  
Qy 337 -----RQIFCOACLRREP-----DSEN 353  
Db 433 RDMTPVVEKYFSNKFNDKNDPYRYIDLSILADGNQIPLGHHKILPESTHSLSPLINDVET 492  
Qy 354 P-----CLISRLMLDAKLYKGARKILHELIFSFFEMEMEKYKLFAMEFVKYKQLOK 406  
Db 493 PPSRYSNTRLOHILYFONRYWKLRKDKQNIWIPTLASSNLYKPIFCQOVVEIFNHIR 552  
Qy 407 E--YISDDHRSISITALS---VOMFTVPTLARHLIEEONVISVITETLEVLPEYL--- 458  
Db 553 SVAYM---DREPOLTAIRECVQVLTCTPCTNAKNIPENQSFQDIWM-SIIDIFKECKVE 607  
Qy 459 -----DRNNKFN-OGYSODKLGRYAVVICDKLYILISK---PTIWERLARMQFEGFR 508  
Db 608 GGVLIQWRVQKSNLTKSYSISFKQGLYTV-----ETLLSKVHDPI--PLRPKEI---I 656  
Qy 509 SPLKILTCMQGMEERROVQOHI-EVDPDWEAAIAIQMOLKNILLMFQEWG-----ACDE 562  
Db 657 SLLTLCKLFGAWKIKRKEGEHVLHEDQNFISYLEVTTTSIYSIQIAEKVSEKSKDSIDS 716  
Qy 563 ELLLVAYKECHKAVMRCSTSFSSSKTVVQSCGHSLETKSYRVSDELVSILHPLSLTAG 622  
Db 717 KLFLNA-----IRIISFLGNRSRLTKLYDSHEVIFKFSVHERVAFMNPLOTMLSF 768  
Qy 623 LHVRLSLRGAVSRHLH---EFVSFEDPOVEVLVEYPLRCLVLVAQVVAEMRRNGLSLISQ 679  
Db 769 LIEKVSLLKDAYEALEDCSDFLKISDFS-----LRSVVLCSQIDVGFWVRNGSVLHQ 820  
Qy 680 VFYYQDVKCREEM--YDKDIIMLOIGA--SLMDPNKFLLLVLQRYELABAFNKTISTKQD 735  
Db 821 ASYI---KNPELGSYSRDIHLNQLALWERDDIPRIIYNILDRWELLDFWTFGEVDYQHT 877  
Qy 736 DLIKOYNTLIEEMQLVLIYIVGERYPVGVNPKTE---EVTREIILHLICIEPMPSAIA 792  
Db 878 VYEDKISFIQOIFAFIYOILTERQYFKTSSLKDRRMDQIKNSIYNLYMKPLSYSKLL 937  
Qy 793 KNLPEN-ENNETGLENVINKVATFKP-GVSGHGVYELKDESKDNMFFYHYSKTOHKS 850  
Db 938 RSPDYLTEDTTFDEALEEVSFVEPKGLADNGVEKLL-ASL-----YAKVDPLK 987  
Qy 851 ASHMK-----KRRKOENKDEALPPPPPEFCFAPFSKVINLNCDIM---MYLTR 898  
Db 988 LNLNEFESSATIIKSHLAKDKDEIAKVVLIPQVS-----IKQLDKDALNLGAFTRNT 1041  
Qy 899 VFERAI-----DTDSNLMTGEMQLMAFHILALG-LLEEKQOLOKA-----PEEVTFD 945  
Db 1042 VFAKVYKLLQVCLDMEDSTFLNELHLVHGIFRDDDELINGKDSIPEAVLSKPTICNLLS 1101  
Qy 946 FTHKASRLGSSAM--NIQMLEKLIKIPQLEGOKDMITWLOMFDVVKRLREKSLIVAT 1003  
Db 1102 IANAKSDVSESIKRVADYLLKEM-----IMKKPNELFESLIA- 1139  
Qy 1004 TSGSESIKNDIETHDKEK-----AERKKAEEAARLHROK IMAOMSALOKNFIEH 1053  
Db 1140 -----SFGNOYVNDYKDKKRGVNLQETEKERRKRLAKHQAARLAKFNNOQTKFMKEH 1194  
Qy 1054 KLMYONTSMPGKESIMEEESTPAVSQYSDYSRIALGPKRGPVSVTEKVELTICILQEQEVEK 1113  
Db 1195 ESEFD---EQDNDVDWV-----GEKVYSEDFTCALQDSS--- 1227  
Qy 1114 IENNAWLSACVQKSTALTQHRKPIELSGEALDP-----LFMPDPLA 1156  
Db 1228 -STDFEIPA-----YHDSPIFRPGNIENPEFMPMDGFFYNDDEKQAVIDDDVL 1277  
Qy 1157 YG-TYTGSCG-----HVMHVCWQKYFEAVQLSSQORIHVDLFDLESYGLCPICK 1206  
Db 1278 EALKENGSCGSKRVFVSCNHHIHNCFKRYVQKKRFS-----SNAPICPLCQ 1324





Db 852 ERRFCGLSTA-----ESLRREIIFKLAGDFTTHSOLYKSLPRDLSSKDELQEVLDVSVYCN 908  
QY 818 PCVSHGVVELKDESLKDFNMVYHYKQHSKAEMQKRRKQENKDKALPPPPPPFC 877  
Db 909 PSGMNGKYSLQSSCKWELDY-----HPRWQSRDLQSAEERF-----SRVC 950  
QY 878 -----PAFSKVINLNCDDIMMYILRTVFERAIDTDSNLWT---EGMLQM 918  
Db 951 GVSALTTLQPRWRMTYPLKGLARIGTKATFOIISALYALQSGTSVKSRAPOGLVIT 1010  
QY 919 AFHILALGELLEKOOLQKAPER---EVTFFDVHKAS-----RLGSSAMNIQMLLEKLAG 969  
Db 1011 ALQLLSLSDICTQORQNSQCCLENSPILELAGLEIIGTAQGEKESLLSLVSLMK 1070  
QY 970 IPQLEGQK-----DMITWT---LQMFDTV-----KRLREKSCILVATTSGSESIKND 1014  
Db 1071 TRMGDGRHQFPPEPGSCNISSIGNLLKKFSAIDSVCMNLLQSLAPVGVQSGFDKVMSSG 1130  
QY 1015 ITHDKEAERKKAAPARLHROKIMQAQMSALQKNIETHKLMYDNTSEMPGKE---DSI 1070  
Db 1131 TSDEKRAKAKERQAA-----IMAKMAEQSKFSLTLSSMDD--DDPRSEFETSDSV 1181  
QY 1071 MEEESTPAVSDYSRIALGP-KRGP-----SVTEKEVLTC 1103  
Db 1182 MEHDEIAVREVCSLCHDPDSKDPVSFLIFLQVGSMTDAICDC 1225

## RESULT 8

S64851

probable membrane protein YLR024c - yeast (*Saccharomyces cerevisiae*)

N:Alternate names: hypothetical protein LI730

C:Species: *Saccharomyces cerevisiae*

C:Date: 01-Aug-1995 #sequence\_revision 24-May-1996 #text\_change 19-Apr-2002

C:Accession: S64851

R:Obermaier, B.; Piravandi, E.; Rinke, M.

submitted to the Protein Sequence Database, May 1996

A:Reference number: S64845

A:Accession: S64851

A:Molecule type: DNA

A:Residues: 1-1872 &lt;OBE&gt;

A:Cross-references: EMBL:Z73196; NID:gl360331; PID:e245500; PID:gl360332; GSPDB:GN00012;

A:Experimental source: strain S288C

C:Genetics:

A:Gene: SGD:UBR2; MIPS:YLR024c

A:Cross-references: SGD:S0004014

A:Map position: 12R

C:Keywords: transmembrane protein

F:189-205/Domain: transmembrane #status predicted &lt;TM1&gt;

F:804-820/Domain: transmembrane #status predicted &lt;TM2&gt;

F:889-905/Domain: transmembrane #status predicted &lt;TM3&gt;

## Query Match

Best Local Similarity 5.9%; Score 535.5; DB 2; Length 1872;

Matches 383; Conservative 320; Mismatches 721; Indels 529; Gaps 93;

QY 104 CGRVFKSGETPSYRCDCAIDPTCVLCMDQFQDSVHKHRY--KMTSTGGGFCDCGDTFA 161

Db 98 CTRLCGFPSETIYCTCTNPLYEICELCFQDKHVNHSYAKVVMRPEGRICHGCD--- 154

QY 162 WKTPGFCVNHPPGRAGTITK-----ENSRCLNPEVIVQARKIPSVIKYVEMPTI 211

Db 155 ----PFAFN-DPSDAFKCKNELNNTPISDNSNVTTDDENVI-----SLLNYLDFLIDIVT 205

QY 212 -WEEKEL-----PPELQIREK-NERYKC-VLFNDEHSHYDIYS----- 249

Db 206 SYKEAEAAHSSERRKASSLMHPNQNSITDDIMEKHECEPLVDNENFVFDNNWNSRKEAH 265

QY 250 LQALDCELAQALHTTAIDKEGRVRAKAGAAQCAEKEDIKSH-----SENVQ-H 301

Db 266 MEWAQIQEEECNVHMDLASTITRLNTPVEYALSIKALEDSHDVTVLQSENFEFD 325

QY 302 PLHVEVLHSEIMAH-----OQFALRLGSMWNKI 329

Db 326 QIAKEFQKENIVVVRKADDIFKRKLTDLTDWLYSLCFKAATSLQNKYALRI--SMLDVM 384  
QY 330 MSYSSDFRQIFCOACLRPEPDSNP-----CLISR----- 359  
Db 385 YSHFSK-----MRVSPNTNPDSKINLLGGFLISNEDSDSWFKPWSLENIEDER 435  
QY 360 ----LMLWDAKL-----YKGAR-----KILHELIFSSP 383  
Db 436 ISKILTYNERLIRAHSPNTVSHFYNYGSRFOYIILNSINILSKSKFKML-KIMASLF 494  
QY 384 FMEWEYKLPAMEFVYKQLOKEYISDDHDSISITALS--OMTVPVTLARHLIE-- 438  
Db 495 SLRDESKFLAAQYIDVYSLYDAVASD-AKEQVTLMSILGQYTFQDPSIANMTISSG 553  
QY 439 --BQNVISVIT-----ETLEVLPEYLDNRNKNFNFOGYS--QDKLGRVAVIC--DLKY 486  
Db 554 FIERTTRFAFTLMAFPEDLMSYLPISL-----YNGFKLPETIRNRTIICFKDLCT 606  
QY 487 IL-----ISKPTIINTERLMQFLEGFRSLFKILTCMQGMEIRROVGOHIEVDP-DWEAAI 541  
Db 607 INSANTVPELLSNEAIFNAIESFSESNVL-----PLKRETKHEVEVENPDFS--- 656  
QY 542 AIQOMQLNILLMQEW-----CACDEELL-----VAYKECHKAVM--RC 579  
Db 657 AFYFFESSILIMTDGYTRISLVKDAFRKQIVLKLDDVAQTRFESLTNSRKAKSPDNA 716  
QY 580 STSFISSSKTVVQS-----CGHSLETKSYRVEDLSVSIHLPLSLTAGHLVRLSLG--- 631  
Db 717 STNENDSNKATLSTVRETICNYVAETINFGVGNVTQYFFNPM--YLFKFKVQWSQCGRYE 775  
QY 632 -----AVSRLHEFVSFEDFOVEVLV-EYPLRCLVLVAQVVAEMRRNGLSISOVFYQD 685  
Db 776 PIPASLTNYINLFEVQDKQKALYSISALSTLVLIQGINVGFWRNGTPTTHQARMYTK 835  
QY 686 VKCREMYDKDIIMLOIGASLMDPNKPLLLVLQRYELAAEFNKTISTKDDOLIKOYNFLI 745  
Db 836 YSMREFTYISDIFNVQFSMACNPDELMTVYLSRWGLKHWAN-GVPMDYDPDTETTVAVV 894  
QY 746 EEMQLVLIYIGE-RVYPGVGNVTK--EVTMR-EIHLILCIEPMPHSAIKNLPENENN 801  
Db 895 NECILLIQLLTVEVRS--VYMSKSGEFTFKSEIHLALCFDTCFAQIVNCIPEHITK 952  
QY 802 ETGLENVINKVATFKKP-GVSGHGVYELKDESLKDFNMVYHYVYKTSKQHSKAEH-----M 854  
Db 953 HPSFDIYLEKANYTSPVSLTDNGIFVLKEKYKDEIDPYIGLSSSSRYDVEKNIRLMA 1012  
QY 855 QKKRRKQENKDEALPPPPPEFCPA-----FSKVINLNCDDIMMYILRTVPERA 903  
Db 1013 NLKKMKYED-----TFVPAKKVKDLLKNTLFSGLYSISSVNTFGLFLKNTLDHI 1061  
QY 904 IDTDSNLWTGMLQMAFHILALGLLEKQOLQKAPEEE---VTFDFYHKASRLGSSAMNI 960  
Db 1062 IKYD-----YDNLPLRVVHLIHLCVNVLNFMFGMLHWEYAIYDTPFCYHS-IGS--ILY 1114  
QY 961 QMLEKLKIGTPOLEGO-KDMITWLQMFDTVKRLREKSCILVATTSGSEST---KNDKI 1015  
Db 1115 YCLLK--DNFSESHGKIREIFRYLM---ETAPHVNVNSYLREQITSTPGILMPTKEDKS 1169  
QY 1016 THDKEAERKKAARLHROKIMQAQMSALQKNIETHKLMYDNTSEMPGKEDSIMBEES 1075  
Db 1170 HRDK---EPERKKHLARLKKLKKLAQOQMFENNSV---DTSDI-----S 1212  
QY 1076 TPAVSDYSRIALGPKRGPVSVEKVL-----CILCQ--EEQEVKI-----ENNA---- 1118  
Db 1213 TPRTTSPS---LSPTRINAENSSNTINSCDDCVCKMPKDDVDFVYFVSQERNICDHG 1269  
QY 1119 -----MWLSACVQKSTALTOHGRKPIELSGEALDPLFMDPDLAYGTYTSCG 1165  
Db 1270 IDFTNPTDVRNINSLFSGKQTKDSIOEN---PQDDGDTRLKFTSCPEVLR-----ACG 1320  
QY 1166 HVMHAYCWQKYFEAVO-LSSQQRTHVDLFDLESEYCLPLCKSLCNTVIP----- 1214  
Db 1321 HGSHTKCLSGHMKSIHQIQNTKNIPL-SYSGSLIYPCVNSLSNSFLPTKNTIDIKRTS 1379



QY 953 -----LGSAMNIQMLEK-----LKGIPOLEGOKDMIT-----WILQMF 987  
Db : : : : :  
1279 LIKFREMETPLVCAMIRLLITETKNGALVVVFGILSGEYDKEVTKGKMIYLARFV 1338  
QY 988 DTVKRLREKSLVATTSGESIKNDITHDKAE-----RRKRAEAAARLHRQKIMAQ 1041  
Db : : : : :  
1339 TILTKLSPVARQIIEGKLREELRISKHSRQEKMAPDPVKKAKEAAKRRMEATM-- 1396  
QY 1042 MSALQKNFIETHKLMYDN--TSEMPGKEDSIMEESTPAVDYSRIALGPKRGPVTEK- 1098  
Db : : : : :  
1397 QNSAKSAQTMKLEKMTGTEMDAEVKNKVDPSQONRKV-----YECPTCGEQNAPNTVENP 1451  
QY 1099 -----EVLTCILCOEBOEVKIENNAVIL-----SACVQKSTALTQHRGKPIELSGEA 1145  
Db : : : : :  
1452 FGLAKLSNFICEEQIDASINTIDLLKFEDEYHVSANLQSETRRRFFSKRRQATFEN 1511  
QY 1146 LDPFLMDPDLAYTYTTCGCGHVMHVCWQRY-----FEAVQLSSQORIHVDLFD---- 1194  
Db : : : : :  
1512 QDIVKVNPLV-CTDLKTCGHAHIAFCFNAYRASLVGSFOEKKLENEKK---KYDGHOR 1566  
QY 1195 -LESGEYLCPLKSLCNTVPI-----IPLQOKINSENADALAAQLLTARWITQTVLAR 1247  
Db : : : : :  
1567 STDRREVGCPCRYTIVNIVPMSFDRPYTPIKTPASPMYSDV-----W-----R 1611  
QY 1248 ISGYNIRHAKGENPIPIFFNOGMGDSITLPHSILSGVESSIKYSN--SIKEMVILPAT 1305  
Db : : : : :  
1612 VMDVLUKAG-----PVFOED-----ERNKYSTNYSTREGGLFELY 1650  
QY 1306 IVRI-----GLKVPDPDERPVPVPM-----L 1325  
Db : : : : :  
1651 IGRIHSADLAERKSSQORCTVSLMVS LAVVWVERSIMKMGVPERKKNQORSMTHELM 1710  
QY 1326 TWSTCA-----FTTOAIENLIGDEGKPLFGALQNRHNGLKALMQFAVAQRITCPQVL 1378  
Db : : : : :  
1711 TASVATSKVDVDFVALSALTNLFAKVTENSEFSRPPSEQPTSKPEAQTEAPN---PEAV 1767  
QY 1379 IQKHILVRLSLVLPNLTKESTPCLLSIDLHV-----LVGAVLAFPSLPDVPDLQ 1430  
Db : : : : :  
1768 VGLSSDEMAAMITKPLRKDTPDCLSKLPALDPKATLVRLMLAVLIDNQSLTKDIQREIA 1827  
QY 1431 PSSVSSYNHLYLFHLITMAHMLQILLT-----VDTGLPLAQ-----VOEDSEEAHSA 1478  
Db : : : : :  
1828 QNMIFASLGVWSSYTLILC-----ILRTGEEKISALNKGEPKTQGLSDHLSQSEAEATCQA 1882  
QY 1479 -----SFFFAEISQYTSIGCDIPGWYLMWSLKNGITPYLRCA-ALFFHYLLGVTP-PEE 1532  
Db : : : : :  
1883 LTYNTEYFKHLAQRLE-SPDVEPTDEVICKTMSCLTIEFLRFSEYELLFHCNLGFNDVNNQ 1941  
QY 1533 LHTNSAEGE-----YSALCSYLSLPTNLLFLFOEYWD-TYRPLLO-----RRCADPALLN 1581  
Db : : : : :  
1942 IHSRNTDLNVLRLVGINAQINVPQK-----ANYTRAPRTQLTNMAKLKRFQPCIVE 1996  
QY 1582 CLKQKNTVYVPRKRNLSLIEPDDYSCLLNQASHFRCPRSADDERKHPVLCLCGATLCS 1641  
Db : : : : :  
1997 PLAWK-----PRR-----ILKPPNTDELFGRYFHREC-NKCSSVPSPYPMCLFCGEIILC- 2045  
QY 1642 QNITCO-----BIVNGEEVGACIFHALHCKA-----RGCAYP 1673  
Db : : : : :  
2046 LNDCCRMHQESGSDRVISNSEIEA---HAEDCSSSGSLFSLVTSMMVVVSRGKQAAIWG 2102  
QY 1674 APYLDYGTDPGLKRGNPLHLSGRERYRKLHLVWQQHCITEIARSOE-----TNQMLFGF 1729  
Db : : : : :  
2103 TIYLDHAKKEEDRLKRGKPLFLCESRLKWLEYDWAEE---QEWQRPQWENMTNSQAFTS 2158  
QY 1730 N 1730  
Db 2159 N 2159

RESULT 10

T48252

eceriiferum3 (CER3) - Arabidopsis thaliana

N:Alternate names: protein TLE22.70

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 20-Apr-2000

C:Accession: T48252

R:Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Bancro

submitted to the Protein Sequence Database, March 2000

A:Reference number: 224489

A:Accession: T48252

A&gt;Status: preliminary

A:Molecule type: DNA

A:Residues: 1-795 &lt;BEV&gt;

A:Cross-references: EMBL:AL162874

A:Experimental source: cultivar Columbia; BAC clone TLE22

C:Genetics:

A:Map position: 5

A:Introns: 203/2; 220/3; 428/1; 488/3; 666/2; 701/2; 723/3; 749/3; 772/3

A:Note: TLE22.70

Query Match 3.2%; Score 292; DB 2; Length 795;  
Best Local Similarity 21.6%; Pred. No. 5.2e-10;  
Matches 180; Conservative 117; Mismatches 265; Indels 270; Gaps 42;  
QY 1038 IMAQMS---ALQKNFIETHKLMY-----DNTSEMPGKEDSIMEESTPAVDYSRIA 1086  
Db : : : : :  
56 LMLQLSDDTISESANMIESIKARLIGNQTEKRSSDGRCKDESME---SLEIAMYQTVR 112  
QY 1087 LGPRGRSVTKEVLTCTILCOEBOEVKIENNAVILSACVQKSTALTQHRGKPIELSG--- 1143  
Db : : : : :  
113 -----NKIENMINQSLTRVDHQPHEAEN-----CSEKNSV-----GGPSTLQGRPP 153  
QY 1144 -----EALDPLFMDPDLAYTYTTCGCGHVMHVCWQRYFEAVQLSSQORI 1188  
Db : : : : :  
154 DIRSROTSSRRPDAGSDGPHIDCD-----GYVLSGCHAVHOSCLERYLKSUKERSGRT 208  
QY 1189 HVD---LFDLESGEYLCPLKSLCNTVPIIP-----LOPQ-KINSENADALAAQLLT 1236  
Db : : : : :  
209 VFEGAHIVDLKAKKEFLCPVCRRLANSVLPECPGDLCSYKSLQDSPTKLRKDALQPSL- 267  
QY 1237 LARWITQTVLARISGYNIRHA-----KGENPIPIFFNOGMGDSITLPHSILSGVESS 1288  
Db : : : : :  
268 ---WLSEALCL-----LRSAAEVIDGDRGKTVTP-----QGDGPRR----- 301  
QY 1289 IKYNSNIKEMVILF-----ATTIYRIGLKVPDPDERPVPMLTWTSTCAFTTOAIENILG 1342  
Db : : : : :  
302 -KDLKSVSKMLWDFYFPKPEDTKRLWL--PPQS-----IVMDTKLYSLISME--IG 350  
QY 1343 DEG-----KPLFGALQNRHNGLKALMQFAVAQR-----ITCPQVLI-QKHLVRL 1386  
Db : : : : :  
351 TRFAKNSMLPYVICDSLYEELKTSKGTLSVLLRVVQSSRTKNTIHRQRFVGMKHLAES 410  
QY 1387 LSVVLPNIKSED-----TPCLLSIDL-----FHVLVGAV--LAFPSLYW 1423  
Db : : : : :  
411 ICYGVSSSSSSSIFGSEGTGSLKNIDLLNWRASDPVLAHDPFSSLMWALFCLPPPLTC 470  
QY 1424 DDPVDLQPSVSSSYNHLYLFHLITMAHMLQILLTVDGLPLAQVQEDSEEAHSSFFA 1483  
Db : : : : :  
471 EESL-----LSLVHIFHSVSLVQTVIAYCACRP-----SELSELNFGENLNLN 512  
QY 1484 EISQYTSGS-----IGCDIPGWYLMWSLKNGITPYLRCAALFFHYLLGVTPPE 1531  
Db : : : : :  
513 DISNALREGGWYFRSNMMDLSCDKD-----TIKYSLPFLRRCALLKWLK--KSTPR 565  
QY 1532 ELHTNSAEGEYSA LCSYLSLPT--NLFLLFQ-----EYWDTVRPLQLQRCADPA 1578  
Db : : : : :  
566 KLIH-----EESDMFELPSDPTDNMDFIYSPQSELNVHVOELKEMFNIPPI-----DII 613  
QY 1579 LLNCLCKOKNTVV-----RYPRKRN-----LIELPDYSCLLNQASHFRCP 1619  
Db : : : : :  
614 LNDELLRSQTQWLQHFQREYRVNRKRSCLITPVVFPOLMKLPNLYQDLQRCIKKRCV 673  
QY 1620 RSADDERKHPVLCLFCGAILCSQ--NICCOEIVNGEEVGACIFHALHCK----- 1666  
Db : : : : :  
674 -NCTKVIEEPVLCCLGS-LCSPISWPCCRE-----SGCPNHAITCAGTGVFLLIRR 724

```
Qy 1667 -----ARCAYPAPYLDEYGETDPLKRGNPLHLSRERYKRLHLVWQOH 1710
Db 725 TTILLQFAROSPWPSPYLDFTGBEDIDMRKRLYLNEERYAALTLYLVGSH 776

RESULT 11
T40238
hypotheical protein SPBC32F12.14 - fission yeast (Schizosaccharomyces pombe) (fragment)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T40238
R:Moreno, S.; Wood, V.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, June 1998
A:Reference number: Z21915
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Accession: T40238
A:Molecule type: DNA
A:Residues: 1-271 <MOR>
A:Cross-references: EMBL:AL023796; PIDN:CAA19375.1; GSPDB:GN000067; SPDB:SPBC32F12.14
A:Genetic:
A:Gene: SPDB:SPBC32F12.14
A:Map position: 2

Query Match 2.4%; Score 218; DB 2; Length 271;
Best Local Similarity 26.0%; Pred. No. 4.1e-06;
Matches 60; Conservative 34; Mismatches 97; Indels 40; Gaps 6;

Qy 65 EKOESVQMSITPLEWY-----LFGEDPDICL-----EKLHSGAFQLC 104
Db 37 ESAKSLNLFVFSALLGYDHTLWTLTPERTIDASFLLRRAQHSGEDVFRHGTCSEKC 96

Qy 105 GRVFKSGETTYSCRDCAIDPTCVLCMDQFQDSVHKHNRHYKMHSTG-GGFCDCGDEAWK 163
Db 97 GHIFRKEVFRCKTSVDSNSALCVKCFRATSHKHDTSTVTSAGSGCCDCGNAAWI 156

Qy 164 TGPCVNHPEGRAGTIENSRCPLNEEVIVQARKIPFSVIKYVEMTWEE--KELPPE 221
Db 157 GDVSKHSHEEDATISNDMIDEIPEKLENSIQTTIDCVLDVLDVDFVSCSPENLKKMPTL 216

Qy 222 LQI--REKNER-----YYCVLFNDEHSHYDHYVYLSQRLAD 255
Db 217 ESILQDEKTSRLSENKYGDIDDSNMYKSLVLWNDEKHSFKFYEQYITTALE 267

RESULT 12
JC5837
364K Golgi complex-associated protein - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 05-Mar-1998 #sequence_revision 13-Mar-1998 #text_change 20-Jun-2000
C:Accession: JC5837
R:Toki, C.; Fujiwara, T.; Sohda, M.; Hong, H.S.; Misumi, Y.; Ikehara, Y.
Cell Struct. Funct. 22, 565-577, 1997
A:Title: Identification and characterization of rat 364-kDa Golgi-associated protein rec
A:Reference number: JC5837; MUID:98093490; PMID:9431462
A:Accession: JC5837
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-3187 <TOK>
A:Cross-references: DBJ:D25543; NID:g516825; PIDN:BAA05026.1; PID:g516826
C:Comment: This protein plays a role in the formation and maintenance of the characteris
C:Superfamily: giantin
F:49-549,624-1176,1238-1707,1763-3114/Domain: coiled-coil leucine zipper #status predict
F:3165-3187/Domain: membrane anchor #status predicted <MAD>

Query Match 2.2%; Score 205.5; DB 2; Length 3187;
Best Local Similarity 20.2%; Pred. No. 0.00098;
Matches 243; Conservative 182; Mismatches 472; Indels 305; Gaps 49;

Qy 130 MDCFQDSVHKHNRHYKMHSTGGGFCDCGDTAWKTGPFVNHPEGRAGTIK-----ENS 183
Db 1057 IDLLQEEITENQATQIKFITGTMDAGDGDGSAVKETSV-----SSPPRAGGGEHWKPELECK 1112
```

---

```
Qy 184 RCPLNEEVIVQARKIPFSVI--KYVEMTWEEKEKLPPELQIREKNERYCYVLFNDEHHS 242
Db 1113 IVDLEKEKTKLQKLEALISRKAILKKAQEKELKEELKEQKDAYRHLQEOFDGOSKE 1172

Qy 243 YDHVYLSQRLDCELAEAQLHTTAIDKGRRAVAKAGAYAAACEAKEDIKSHSENVSO-- 300
Db 1173 NENIRAPLQLOAKESTDOOLPGT-----GQEPTHGSEGLSLEGTEPASESLHAAQPS 1227

Qy 301 HPLHVEVLHSEIMAHQ-----KPALRLGSMNMKIMSYSSDFRQIFCQACLREEP 349
Db 1228 HPGETATLQATVSAQIQDLQKEIEVEKELEL-----KISSTTSELTKKSEEVLLLOEQ 1282

Qy 350 DSENPCLISRLMLNDKLYGARKILHILIFSF--PMEMEYKLFAMEFVYKYQLQKEY 408
Db 1283 INEOGLEIQNLKAASHEAKAHTEQKQOLESQKLIADLHLKTLQPEL---ETLQKHV 1338

Qy 409 ISDDHRSISITALSVMQMETVPTLARHLIEEQNVISVITETLLEVLPEYLDNRNKNFQG 468
Db 1339 GQKEBEVSYLVGOLGEKEQTTLTVQTEM--EEQ-----ERLKAHTOLEMQAKEH--- 1387

Qy 469 YSQDKLGRVYAVICDLK-----YILISKPTIWTIRLRMQFLEGFRFLKILTCMQMEEI 523
Db 1388 --EERLKQVQVEICELKQPKPELEESKAKQLOKLQAALISRKALKENKSLQEQLS 1445

Qy 524 RRQVQGHIEVD--PDWEAAIAIQMLKNILL----MFQEWACADEELLLVAYKECHKAVMR 578
Db 1446 ARDAVEHLTKSLADVESQVSNQEKDALLGLALLQE---ERDKLIV---EMDKSLLE 1498

Qy 579 CSTSFISSTKTVVSCGHSHLETKSVRSVSDLYSIH--LPLSRTLAGLHVRLSR--LGAVS 634
Db 1499 -NQSJGGSCESLKLALGGLTDEKE--KLMKELSVRCSKIAETSEWQEKHKELOKEYEVLL 1556

Qy 635 RIHEEVSFDFQVEVLVEYPLRCLVLVAQVAAEMW-----RRNGLSLISQVFF 682
Db 1557 QSYENVSNAEARIQHVE-----SVRQEKQEVYAKLSAESDKRERKQLQDAEQEME 1609

Qy 683 YQDVCRREMYDKDIIMLOJ-----GA-----SLMPDNKFLLLVLORYEL 722
Db 1610 EMKEMRKPAKSKQKQKILEEBEENDRLRAEAQPVGGANESMEALLSNASLKLELERITL 1669

Qy 723 AEAFNKTISTKDODLKOYNTLIEEMLOVLVIYIGERYVPGVGNVKEEVTMEIHLIC 782
Db 1670 E--YKTLSEFEALMAERTLSEETRNKLLQVEAQELQASLETTEKSEDEPKDV---- 1721

Qy 783 IEPMPSHIAKNLPENENNTEG--LENV--INKVATFKPGVSGHGVYELADESLKDFNMVF 840
Db 1722 IEVTEAVVGKSEQDLSLENAKLEDAEATLLANSAPGVSE-----TFSSHDDINNVL 1775

Qy 841 YHYKTOHSAE--HMOKKRRKQ-----ENKDEALPPPPPEFCFAPSKV-----IN 885
Db 1776 QOOLDOLKGRIAELEMEKQKDRLELSQTLNEKNAL----LTQISAKDSSELKLEEEVAKIN 1831

Qy 886 LLNCIDIMVILRT--VFERRAIDTDSNLWTEGLMAFHILALG----- 926
Db 1832 MLNQOIQEELSRVTKUETAEKEDLEERLMMQLAELNGISIGNYQDVDTDAQIKNEQLE 1891

Qy 927 -----ILLEEKQOLQKAP-----EEVTFDFYHK-----ASRLGSSAMNTQML 963
Db 1892 SEMQNLKRCVSELEEEKQOLVKEKTKVSEIRKEYMEKIQGAKQPGSKIHAKELQ--ELL 1950

Qy 964 LEKLKIPOLEGOKDMTW---ILQMFDTVVKRL----- 993
Db 1951 KEKQEVKQL--QKDCIRYLGRISALEKTVKALEFVHTESOKDLDATKGNLAQAVEHHKK 2008

Qy 994 -----REKSLIVATT-----SGSESIKND-----EI 1015
Db 2009 AQAELSSEFKILLDDTQSEARVLADNLKLLKELQSNKESIKSQIKQKDEDLRLLEQAE 2068

Qy 1016 THDKKAERKRAEARLHRKQIMAO-----MSALQKNFIETHKLMYDN 1059
Db 2069 KHRKEKKNNQEKLDA--LHREKAHVEDTTLAEIQVSLTRKDKMKELQOQSLDSTLAQAAF 2126
```



```
QY 1060 TSEPGKEDSIMEESTPAVSDYSRIALGPKR-----GPSVTEKEVLTCLCOEEQEVKI 1114
Db 2127 TKSMESLQD-----DRDRVIDEAKRWEQRFDAIQTK-----EEVRLKE 2166
QY 1115 EN 1116
Db 2167 EN 2168

RESULT 13
A56539
N:Alternate names: macrogolin
C:Species: Homo sapiens (man)
C:Date: 19-Oct-1995 #sequence_revision 26-Jan-1996 #text_change 10-Dec-1999
C:Accession: A56539; S37536
R:Seelig, H.P.; Schranz, P.; Schroeter, H.; Wiemann, C.; Griffiths, G.; Renz, M.
Mol. Cell. Biol. 14, 2564-2576, 1994
A:Title: Molecular genetic analyses of a 376-kilodalton Golgi complex membrane protein
A:Reference number: A56539; MUID:94187728; PMID:7511208
A:Accession: A56539
A:Molecule type: mRNA
A:Residues: 1-3259 <SEE>
A:Cross-references: EMBL:X75304; NID:g405714; PIDN:CAA53052.1; PID:g405715
C:Genetics:
A:Gene: GDB:COLGB1; GCP: GCP371
A:Cross-references: GDB:454958
A:Map position: 3q13.31-3q13.31
C:Superfamily: giantin
C:Keywords: coiled coil; Golgi apparatus; transmembrane protein
F:3238-3254/Domain: transmembrane #status predicted <TM>

Query Match 2.1%; Score 190.5; DB 1; Length 3259;
Best Local Similarity 18.0%; Pred. No. 0.0086;
Matches 221; Conservative 207; Mismatches 424; Indels 377; Gaps 49;

QY 169 VNHEPGRAGTIKENSRCPLNEBVIYQARKIPFSVTKYVVMWIMEEKELPPELQ---IRE 226
Db 12 VLHLEGDDDDTDDNRRAPDPELHQESDMEFNNTTQEDYQERLAYAE-QLVVVELKDIIRQ 70
QY 227 KNERYCYVLFNDEHSHSVHVIYSQALDCELAELQHT---TADK--EGGRAVKAGA 280
Db 71 KDVL-----QKDEALQERKAADNKIKKLHAKAKLTSUNKYIEMKAGGVV 121
QY 281 YAAOQAEKEDIKSHENYSQHLVHEVLHSEIMAHQKFAIRLGSWMNKIMSYSSDRQIF 340
Db 122 LPTEPQSEQLSKHDKSSTEREMEIEKIKHKIQEKEELSTL----- 163
QY 341 COACLREPDSENCLISRLMLWDALYKGAARKILHELIFSSFFMEMEYKKLFAMEFVKY 400
Db 164 -QAQL-TQAARQPAQSS-----TEME-----EFVMM 188
QY 401 YKOLQKEYISDDHDSISITALSVMFTVPTLARHLIEQN-----VLSVIT 447
Db 199 KQQLQEK-----EEFISTLAQLSQTQAQAAQVVRKDAFETQVRLHEDELLQVIT 242
QY 448 ETLLEVLPEYLDNRNKNFNGQYSQDKGRVYAVICDLKYLITS---KPTIWTERLHM-- 501
Db 243 QADVETEQQKRLVLRQKLEHEESILGR--AQVVDLLOQELTAEBQRNQLISQQLQOME 300
QY 502 -----QFLEGFRSFLKILTCMQGMEETIRQVGQHIEYDPDWEAAIAIQMKNLILMF 554
Db 301 AEHNTLRNVTETERESKILLEKMEVAERKLSFH-----NLQEEMHHLLEQF 349
QY 555 QEWACADEEL--LIVAYECHKAVMRCSTSF--SSSKT--VVQSCGHSLETYSRYRVSDEL 609
Db 350 EQAGQAQAELESRYSALEQKHKAEMEKTSHLSLQKTQELQSACDALKDQNSKLLQDK 409
QY 610 VSIHPLSRTLAGLHVRLSRLGA-----VSRL-----HEF---VSPEDFOVEVLVEYPLRC 657
Db 410 NEQAVQSAQTIOQLEDQLOQSKSEISQFLNRLPLQHQEATASQSPFDVYNEGTQAVTEEN 469
QY 658 LVLVAQVVAEMRRNGLSLSISQVFFYQDVKCREMYDKDIIMLQI-----GAS 705
```

```
Db 470 IASLQKRVVELENEKGALLSSI--ELEELKAENEKLSQITLLEAQNRTGADREVSVIS 528
QY 706 LMD-PNK-----FULLVLQRYELAE 725
Db 529 IVDIANKRSSSAEESGQDVLENTFSQHKKELSVLLEMKAEQAEIATLKLQLOGRABEA 588
QY 726 FNKTISTKQDQLIKQY--NTLIBEMLOVLIYICERY----- 760
Db 589 DHEVL---DQEMKQMEGEGIAPIKMKVFLDTQODFPLMPNESSLPAVEKEQAOSTBHQ 645
QY 761 -----VPGVGNVTKEEVTMRTEIHLHLCIEPMPHSA- 790
Db 646 SRTSEBESISLDAGVELKSTKQDGDKSLSAVPDICOQHODELE-RLKSOILELELNFHKAQ 704
QY 791 --TAKNLPENNETGLNVINKVATFKKPGVSGHGYV---ELKDESL---KDFNMYF- 840
Db 705 EIVEKNLDEKAKETSINLQLEIE--PKKNADNNSAFTALSEERDQLLSQVKELSMYTE 761
QY 841 --YHYSKTQHSKAEHMQKRRKQENK---DEALPPPPPPPCFPAPSKFVINLLNCDIMMYI 895
Db 762 LRAQVKQLENNLAEAEQRRLDYESQTAHNLLE-----QIHSLSIEAKSKDVKTEV 814
QY 896 LRTVFE-----RAITDTSNLWTEGMLQMAFH 921
Db 815 LQNELDDVQLQFSEQSTLIRSLQSLQNKSEVLEGAERVHRHISKEVELSQALSQKELE 874
QY 922 ILALG--LLEK---QOLQKAPPE-----EVTDFYHKASRLG---SSAMTQMLLEK 966
Db 875 ITKMDQLLLEKKRDVETLQOTIEEKDOQVTEISFMTKMKVQNLNEEFSLQVEIKTLKEQ 934
QY 967 LKGI-----PQLEGQKDMITWILQMFDTVKRLREKSLIVATTSGESIKNDEITHD 1018
Db 935 LNLLSRAEEAKKEQVEEDNEVSSGLKQNYDEM-----SPAGOISKEELQHE 980
QY 1019 -----REKAPRKKAELHRRQKIMAOMSALQKNFIETHKLMYDNTSEMPEKEDSIMEE 1073
Db 981 FDLKKENEQRRKKLAALINRKELLQVRSLER---ELANLKDESKKEIPLSETERGEV 1037
QY 1074 ESTPAVSDYSRIALGPKRGPSTVEKEVLTCLCOE-----EGBVKIENNAWVLS- 1122
Db 1038 EEDKENKEYSE-----KCVTSK-----CQEIYILYKQTISEKEVELQHIKRDLEE 1082
QY 1123 --ACVQKSTALTQHRKPIELSGEALDPL 1149
Db 1083 KLAEEQFOALVKOMNQTLDKTNQIDLL 1111

RESULT 14
A57013
N:Alternate names: endosome-associated protein
C:Species: Homo sapiens (man)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: A57013; S44243
R:Mu, F.T.; Callaghan, J.M.; Steele-Mortimer, O.; Stenmark, H.; Parton, R.G.; Campbell, J. Biol. Chem. 270, 13503-13511, 1995
A:Title: EEAL, an early endosome-associated protein. EEAL is a conserved alpha-helica
A:Reference number: A57013; MUID:95286647; PMID:7768953
A:Accession: A57013
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1410 <RES>
A:Cross-references: GB:L40157; NID:g1016367; PIDN:AAA79121.1; PID:g1016368
R:Seelig, H.P.
submitted to the EMBL Data Library, April 1994
A:Reference number: S44243
A:Accession: S44243
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-254,'C',256-257,'A',260-276,'A',278-283,'A',285-519,'D',521-574,'EQ',
A:Cross-references: EMBL:X78998; NID:g475933; PIDN:CAA55632.1; PID:g475934
C:Genetics:
```

A; Gene: GDB:EEA1  
A; Cross-references: GDB:136996  
C; Superfamily: human early endosome antigen 1  
C; Keywords: calmodulin binding; endocytosis; metal binding; peripheral membrane protein;

Query Match 2.0%; Score 186; DB 1; Length 1410;  
Best Local Similarity 18.5%; Pred. No. 0.0047;  
Matches 204; Conservative 200; Mismatches 424; Indels 272; Gaps 45;

**y**            181 ENSRCPNEEVQARKIFPSVIKYVVENTIWEEEKLPPELQIREKNERYCVLFNDEH 240  
       | : : | | : : | : : | : : | : : | : : | : : | : :  
**b**            275 ELSKGPQEVRYVQELQKLSSVNELTQNQTLTENLLKKEDDYTKLEKH -----NEES 329

```

241 HSYDHYIYSL-QRALDCELAELUHTTADKEGRAVKAGAYAACQEAKEIKSHSENVS 299
      |  ::  : | : ||| : ::|
330 VSKNKIQTATLHQKDLDCQQLQSLR-----SASE-TS 359

```

a y  
300 QHPLHVEVLHSEIMAHQKFAIRLGSGWMNKIMSYSDFRQTFCACUREEPSDNPCLISR 359  
| : | | | : | | | | : | : | : | : |  
b b  
360 LHRIHVE-LSEGEATQKLKEELSEVETRYQHLEAFKQLQQQ---REEKEQHGLQLOQE 415

```

360 LMLWDKLYKGARKILHELIFSSFFMEMEYKKLFAMEFVKYKQLOKEYISDDHDRSTSI 419
      : ||          | | : | : | | | | | : :
416 INQLHSKL-----LETERQLGEAHGRLKEQRQLSSEKLMDEQQ---V 455

```

**a**

TALSVMQETVPTLARHLTEEQNVSITETLLEVLPEYLDNRNKNF--FQGYSQDKLRGV 477  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
**b**

ADQLK-----LSR--LEEQLKEKVNSTELQ----HQLDKTKQHQQEQALQQSTAKL 504

```

478 YAVICDLKYLILSKPTIINTERLWQFLEGRSFLKILTQMGMEETRRQVGHIIEVDPDW 537
      ||::|| - - - : :: || - - - : :: || - - - : :: || - - - : :: |
505 REAQNDLEQVL---RQIGEKDQKIQNLEAL--LQSKENISLLEKEREDLYAKIQAG-EG 558

```

538 EAAIAIQMLKNTLLMFQEWACDELLV-----AYKECH-----KAVMRC 579  
 559 ETAVLNLOLEKNTLQ-----DEVTLTNVKNQSESHKQAQENLHDQVQEQKAHRA 611

```

580 STSFISSKTVVQSCGHSLETKSYRVSEDLVSI-----HLPL 619
      : : : : : : : : : : : : : : : : : : : : : : : :
612 AQDRVLSLETSNELNSQLNESKEKVSQLDIQIKAKTELLLSAAAKTAQRADLQNHLDY 671

```

617 SR-TLAGHVRSRLGA-----VSRLEH---FVSFE-----DFQ 640  
672 AQNALQDKHQELNKITTQDQVAKLQDKQHCQSLESHLKEYKEYLSLEQKTEEEQC 731

b

732 IKKLEADSLEV KASKEQALDLOOQRQLN-----TDLELRATELSKQLEMEKEIVSS 783  
:: | : | : : : | : : | : | : | : :  
647 VEVLVEXPLRCLVLVAQQVAAEMMRNRGSLISQVFYQDVKRCEEMYDKDIIMLQ--IGA 704

784 TRLDLQ-----KSEALSIKQLTKQEEKILKQDFLTSQE-TKIQHEELNNRTQ 835

836 TTVTELQVKVMEKEALMTLSTVKDKLSKVSDSL-KNSKSEFEKENQKGAALDLDETC 894

014  
015  
016  
017  
018  
019  
020  
021  
022  
023  
024  
025  
026  
027  
028  
029  
030  
031  
032  
033  
034  
035  
036  
037  
038  
039  
040  
041  
042  
043  
044  
045  
046  
047  
048  
049  
050  
051  
052  
053  
054  
055  
056  
057  
058  
059  
060  
061  
062  
063  
064  
065  
066  
067  
068  
069  
070  
071  
072  
073  
074  
075  
076  
077  
078  
079  
080  
081  
082  
083  
084  
085  
086  
087  
088  
089  
090  
091  
092  
093  
094  
095  
096  
097  
098  
099  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832

955 QQLQGNINEL-----KQSEKQKKQIEALQGELKIAVLQKTELENKLOQQ---LTQAAQ 1005

1006 ELAAEKEKISVLQNNYKESQETFKQLQSDPYGRESELLATRODLKSVEEKLS-----IAQ 1066

1061 EDLISNRNQIGNKLIQELTKATATLEQDSAKKEOQLQERCKALQDIQKESLUKEKELV 1120

Qy	1017	HDK-----EKAERKKAEARLUHQKIMQAQSALOK--NFETHKLMDYNTSEMPGCKD	1068
Dd	1121	NESKLAIEIEIKCRQEKEITTKLNEELSHKLESIKEITNLKDAKOLLIQOKEELOQKAD	1180
Qy	1069	SI---MEEESTPAVSDYSRTALGPCKRGPSYTEKEVLTCILCPEEQEVK----IENNAWVLS	1122
Dd	1181	SLRAAAVEQE-----KRNOQLKDQVK-----KEEELKKEFIEKEAKLHS	1220
Qy	1123	ACVQKSTALTQHRCKPIELS	1142
Dd	1221	EIREKEYGMKKHENEANEAKLT	1240

RESULT 15  
T08621  
centrosome associated protein CEP250 - human  
C:Species: Homo sapiens (man)  
C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000  
C:Accession: T08621  
R: Mack, G.J.; Rees, J.; Sandblom, O.; Balczon, R.; Fritzler, M.J.; Rattner, J.B.  
Arthritis Rheum. 41, 551-558, 1998  
A: Title: Autoantibodies to a group of centrosomal proteins in human autoimmune  
A: Reference number: Z16462; MUID:98165428; PMID:9506584  
A: Accession: T08621  
A: Status: preliminary; translated from GB/EMBL/DDBJ  
A: Molecule type: mRNA  
A: Residues: 1-2442 <NAC>  
A: Cross-references: EMBL:AF022655; NID:q2832236; PIDN:AAC06349.1; PID:q2832237  
A: Experimental source: cell line HeLa

Query Match 2.0%; Score 182; DB 2; Length 2442;  
Best Local Similarity 20.9%; Pred. No. 0.019;  
Matches 256; Conservative 187; Mismatches 478; Indels 304; Gaps 60;

QY	212	WE---EEXELPELOIR	REKNERYCYVLFENDEHSHSDVHVIYSQALDCEAE-----AQ	262
DB	88	WENVEPNL-DEL	VLVRLEEEQRCESLAE-----VNTQRLRHEKADVVNKALRAD	137
QY	263	LHTTAIDKEGRRAV	KAGAYACAOEAKEDIKSHSENVSQHPLHVLEVLHSEIMAHQKFAURL	322
DB	138	VEKUTVDWSARD	MLMKESQWQMEQFFKGYLKGEGRLLS---LWREVVTFRHIFLEM	194
QY	323	GSMWNK-IMSYSSD	FRQI---FCOACLR-----EEP-----DSENPCILSRMLWDA	365
DB	195	KSATDRDLMEL	KAHVRLSGSLTTCCLRTVGAQSRPNVGRMDGREPAQLLLLLLAKTQ	254
QY	366	KLYGARKILHELIF	SFFFMEMYKKL--FAMEFVKYKQLOKEYISDDHRSISITALS	423
DB	255	ELEKEAHERSQELI	QKLSQSGDLKAEALQDRVTELSSALLTOSOKO--NEDYEKMIKALRET	312
QY	424	VOMETV-----	PTLARHLIE---QNVISVITETLEV-----LPEYLDNRN	461
DB	313	VEILETHNTELM	HEASLSRNAQEKLSLOVYKIDITQVWVEEGDNTAQSGGLENSELE	372
QY	462	---NKTNFQOYSODK	-----LGRVYAVICDLKY-----ILSKPTIWTPE--	497
DB	373	SSIESQFDYQ--	DADKALTLVRSVLTTRRQAVQDLRQLAGCOEAVNLLQQQHDOWEEG	430
QY	498	---RLRMOFEGFRS	FTKILTC-MQG-----MEEIRVOVGQHIEV--DPDWE-A	539
DB	431	KALRQRLOKLT	GERDTLAGOTVDLQGEVDSLKSERELLOKARELROQLEVEQEAURLR	490
QY	540	AIAOTMQLKNILL	FOEWCACDEBLLLVAYKECHKAVMRCSTSPISSKTVQVSCGHSLE	599
DB	491	RVNVELOQDGS	AQGCK-----EE---QOEELHLAV-----RERERLOEMLMGLE	532
QY	600	TKS-----YRV	SEDLVSIHLP---LSRRTLAGLRVRLSGVARSRLHEFVSFED-FQVE	648
DB	533	AQSESISELIT	LEALESHLEGEELLROQEVTEAALAR--AEQSTAELSSSENTLKTE	590
QY	649	VLVEYPLRCLV	LVAVQVAMWRNRNGLSLISOVFYV-----QDVKCREMYDKDIIIMIGI-	703
DB	591	VA---DLRAAAVK	LSALNEALADPKVGNLQOLLOLEBENQSVCSMRMAEAQARNALQVDL	647

```

QY 704 -----ASIMDPNKELLVLQRYELA-----EAFNKTIKQODLIKQYN----- 742
Db 648 ABAEKREALWEKNTHLEAQLQAEAGAEALQADLRDIOEKEEIOKKISESRHQEPAAT 707
QY 743 TLIIEMLQ-----VLIIVIGERYVPGVGNVTKKEEVTWREIHLICIE-----PMP 787
Db 708 TOLEQLHQEAKRQEEVLARAVQEKEA-----LVREKALE--VRLQAVDRDRLAAQLQ 760
QY 788 HSAIAKNLPENNETGLEN-VIN-----KVATFKPGVSGHGVYELKDESLKDF 836
Db 761 GLSSAKELLESSLFEAQOQNSVIDEPQGLEVQIQVTOAKEVIOGE-VRCLE----- 814
QY 837 NMYFYHYSTQSKAEHMKRKRKQENKDEALPPPPPEFCPAFSKVINLLNCDIMYIL 896
Db 815 -----LPTERSQAQERDAARQAQAEQCKTALQEQKAAHEKVNQ-----L 858
QY 897 RTVFERAIDTDSNLTWTEGLMAFHILALGLLEEKQLOKAPKEEVTDFYHKASRLGSS 956
Db 859 REKWEK-----ERSWHQOELAKALESLEREKMELEMLRLEKQOQTEMEAIQAREERTQAE 913
QY 957 AMNIQMLEKLGIPQLEGQKDMITWILQMFDTVKRLREKSCLI--VATTSGSESINKDE 1014
Db 914 SALCOM-----QLETEKERVSVLETLITQKELADASQQLERLRQDMKVQKLKEQE 964
QY 1015 IT-----HDKEKAERKKAARLHROKIMA---OMSALQKNFTIETHKLMYDNTSEMPGKE 1067
Db 965 TTGILQTOLEAQARELK-EAARQHRRDLAALQOESSLLQDKMDLQKQVEDLKSQVAD 1023
QY 1068 DS--IME--EESTPAVSYSRIALGPKRGPSVTEKEVLTICLQEEQEVKIENNAMVLS 1122
Db 1024 DSQRLVEQVQKLBETQYNNRIQELER-----EKASITLSLMEKEQRLVLQOED--S 1076
QY 1123 ACVQKSTALTQH-----RGKPIELSGE-----ALDPLFMDPDLAYGTYTSGCGHV 1167
Db 1077 IRQOELSLALQDMQEAQGEQKELSAQOMELLRQEVKEKEADFLAQEAQLLEEL--EASHI 1133
QY 1168 ----MHAVCWQKYFEAVQLSSQORIHVDLFDLESGEYLCPLCKSLCNTVPIPIPOFOKI 1223
Db 1134 TEQQLRASLWAQEAQAAQL-----HLRLRSTES-----QLEALAAEQ 1171
QY 1224 NSENADALAOILLTARWIOTVLARI 1248
Db 1172 PCNQAQAQLASLYSALQOALGSV 1196

```

Search completed: September 25, 2003, 14:56:30  
Job time : 78 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 25, 2003, 14:46:05 ; search time 33 seconds  
(without alignments)  
2471.040 Million cell updates/sec

Title: US-09-724-126a-19

Perfect score: 9141

Sequence: 1 AMEGNNADEEAGGTERMEIS.....ETARQETNQMLFGFNQQL 1734

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	990	10.8	1958	1	UBR1_SCHPO	O60152 schizosacch
2	866	9.5	2052	1	UBR8_SCHPO	O13731 schizosacch
3	691.5	7.6	1950	1	UBR1_YEAST	P19812 saccharomyc
4	628.5	6.9	1941	1	UBR1_KLULA	O60014 kluyveromyc
5	208	2.3	2230	1	G034_HUMAN	Q13439 homo sapien
6	190.5	2.1	3259	1	GIAN_HUMAN	Q14789 homo sapien
7	189.5	2.1	8797	1	SNE1_HUMAN	Q8nf91 homo sapien
8	182.5	2.0	2442	1	CEP2_HUMAN	Q9bv73 homo sapien
9	177	1.9	6885	1	SNE2_HUMAN	Q8wxh0 homo sapien
10	176	1.9	1829	1	MY3A_CHICK	Q02440 gallus gall
11	175.5	1.9	1427	1	REST_HUMAN	P30622 homo sapien
12	175.5	1.9	1526	1	MY52_SCHPO	Q9us16 schizosacch
13	175.5	1.9	2663	1	CENE_HUMAN	Q02224 homo sapien
14	170	1.9	2298	1	CU05_HUMAN	Q9y3r5 homo sapien
15	168.5	1.8	3911	1	AKA9_HUMAN	Q99996 h a-kinase
16	166	1.8	1163	1	SBCC_CLOAB	Q97fk1 clostridium
17	164	1.8	2096	1	BP28_DROME	O9vm75 drosophila
18	162	1.8	2335	1	TOR1_SCHPO	O14356 schizosacch
19	158	1.7	2033	1	EVPL_HUMAN	Q02817 homo sapien
20	157.5	1.7	3214	1	BPAL_HUMAN	Q03001 homo sapien
21	156.5	1.7	1790	1	USO1_YEAST	P25386 saccharomyc
22	156.5	1.7	2710	1	TOXA_CLODI	P16154 clostridium
23	156.5	1.7	2869	1	RBPL_PLAVB	Q00798 plasmodium
24	156	1.7	1853	1	MY5A_MOUSE	Q99104 mus musculus
25	155.5	1.7	1216	1	P1B1_RAT	P10687 rattus norv
26	155.5	1.7	8545	1	ANCL_CAEEL	Q9n4m4 caenorhabdi
27	154	1.7	2469	1	TEDG_HSVSA	Q01056 herpesviru
28	154	1.7	3685	1	MDU_HUMAN	P11532 homo sapien
29	153.5	1.7	1875	1	MLP1_YEAST	Q02455 saccharomyc
30	153	1.7	1682	1	MSP1_PLAF3	P15998 plasmodium
31	152.5	1.7	1325	1	G160_MOUSE	P55937 mus musculus
32	152.5	1.7	3433	1	UTRO_HUMAN	P46939 homo sapien
33	151.5	1.7	1855	1	MY5A_HUMAN	Q9y411 homo sapien

#### ALIGNMENTS

##### RESULT 1

```

UBR1_SCHPO
ID  UBR1_SCHPO          STANDARD;          PRT;  1958 AA.
AC  O60152;1998 (Rel. 37, Created)
DT  15-DEC-1998 (Rel. 37, Last sequence update)
DT  15-DEC-1998 (Rel. 37, Last annotation update)
DT  28-FEB-2003 (Rel. 41, Last annotation update)
DE  N-end-recognizing protein (Ubiquitin-protein ligase E3 component) (N-
DE  recognin).
GN  UBR1 OR SPBC19C7.02 OR SPBC32F12.14.
OS  Schizosaccharomyces pombe (Fission yeast).
OC  Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC  Schizosaccharomycetales; Schizosaccharomycetaceae;
OC  Schizosaccharomycetes.
OX  NCBI_TaxID=4896;
RN  [1]
RP  SEQUENCE FROM N.A.
RX  MEDLINE=21559218; PubMed=11702950;
RA  Kitamura K., Katayama S., Dhut S., Sato M., Watanabe Y., Yamamoto M.,
RA  Toda T.;
RT  "Phosphorylation of Mei2 and Stell by Pat1 kinase inhibits sexual
RT  differentiation via ubiquitin proteolysis and 14-3-3 protein in
RT  fission yeast.";
RL  Dev. Cell 1:389-399(2001).
[2]
SEQUENCE FROM N.A.
RC  STRAIN=972;
RX  MEDLINE=21848401; PubMed=11859360;
RA  Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA  Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA  Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA  Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA  Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA  Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA  James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA  Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA  Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA  Rutherford J., Rutter S., Saunders D., Seeger K., Sharp S.,
RA  Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA  Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA  Woodward J., Volkart G., Aert R., Robben J., Grymonprez B.,
RA  Weltjens I., Vansteels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA  Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA  Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA  Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA  Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA  Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA  Lucas M., Rochet M., Galliard C., Tallada V.A., Garzon A., Thode G.,
RA  Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA  Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA  Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA  Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT  "The genome sequence of Schizosaccharomyces pombe.";
RL  Nature 415:871-880(2002).
CC  -!- FUNCTION: RECOGNITION COMPONENT OF THE N-END RULE PATHWAY. BINDS
CC  TO PROTEINS BEARING AMINO-TERMINAL RESIDUES THAT ARE DESTABILIZING

```

```

34  151      1.7      3660  1  DMD_CHICK
35  151      1.7      4870  1  RYR3_HUMAN
36  151      1.7      5430  1  MACF_HUMAN
37  150.5    1.6      1539  1  Y373_HUMAN
38  150.5    1.6      5327  1  MACF_MOUSE
39  150      1.6      5938  1  MAC4_HUMAN
40  149.5    1.6      978   1  RA50_AQUAE
41  149.5    1.6      1453  1  Y373_BOVIN
42  149.5    1.6      1679  1  Y109_YEAST
43  149.5    1.6      2017  1  MYSN_DROME
44  149.5    1.6      2273  1  HFAL1_YEAST
45  149      1.6      1727  1  ALM1_SCHPO

```

```

P11533 gallus gall
Q15413 homo sapien
Q9upn3 homo sapien
Q15078 homo sapien
Q96pk2 mus musculi
Q96pk2 homo sapien
O67124 aquifex aeo
Q9tu23 bos taurus
P40457 saccharomyc
Q99323 drosophila
P32874 saccharomyc
Q9utk5 schizosacch

```











QY 1586 KNTVRYPRKRN-----SLELPDDYSCLLNQASHRCPRSD-----DERKHPV 1630  
 Db 1642 QESIVSHPYLENIPYECGIIKIDLSKYLNTYVYQSKELKRLREERSOHMKNAONRDLFK 1701  
 QY 1631 LCLFCGAIL-----CSQNICCCQIVNGEEVGCIFHALHCKARGCAYPA 1674  
 Db 1702 ICLTCGVKVLRADRHETKHLNKNCFKPGCAFMPNSSEV-CLH-LTQPPSNIFISA 1757  
 QY 1675 PYLDEYGETD-PGLKRNPHLSRERYKHLVWQHCIIIEEIAQSE-----NQM 1725  
 Db 1758 PYLNSHGVEGRNMRGDLATLNLKRYEHLNRLWINNEIPGYSIKRVMGDEPRVILSNFG 1817  
 QY 1726 LFGFN 1730  
 Db 1818 LFAFN 1822

RESULT 4  
 UBRL\_KLUJA STANDARD; PRT; 1941 AA.  
 AC O60014;  
 DT 15-DEC-1998 (Rel. 37, Last Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE N-end-recognizing protein (Ubiquitin-protein ligase B3 component) (N-recognin).  
 GN UBRL.  
 OS Kluyveromyces lactis (Yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Kluyveromyces.  
 OX NCBI\_TaxID=28985;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Waller P.R.H., Varshavsky A.;  
 RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: RECOGNITION COMPONENT OF THE N-END RULE PATHWAY. BINDS  
 CC TO PROTEINS BEARING AMINO-TERMINAL RESIDUES THAT ARE DESTABILIZING  
 CC ACCORDING TO THE N-END RULE, BUT DOES NOT BIND TO OTHERWISE  
 CC IDENTICAL PROTEINS BEARING STABILIZING AMINO-TERMINAL RESIDUES.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF061554; AAC15841.1;  
 DR PIR; T30554;  
 DR InterPro; IPR003126; Znf\_Nrecognin.  
 DR Pfam; PF02207; zf-UBR1; 1.  
 DR SMART; SM00184; RING; 1.  
 DR SMART; SM00396; Znf\_UBR1; 1.  
 KW Ligase; Ubl conjugation pathway.  
 SQ SEQUENCE 1941 AA; 223682 MW; 37C2E1BCA0803268 CRC64;

Query Match 6.98; Score 628.5; DB 1; Length 1941;

Best Local Similarity 20.4%; Pred. No. 2.7e-30;

Matches 406; Conservative 315; Mismatches 683; Indels 587; Gaps 92;

QY 40 FYTAFLLHLAQLVPEIYFAEMDPLEKEESVQMSIETPLEWYLFGEPPDICEKLKHS 99  
 Db 51 YYYNIMSDSGRLPHMTATNREFFRNVOQAMEIKLSKFWYKIDENG--HSGFNAG 108  
 QY 100 AFQLCGVFKSGSETTYSCRCADIPTCVLCMDCQFQDSVHKNHRY--RMHTSTGGGFCDCG 157  
 Db 109 --RICGAKFRGVEPIYRKCESFDDTCVLCVNCFNPKDVGHHVYTSICTEFNGICDCG 166  
 QY 158 DTEAWTKGPCVNVHEPGRAGTIKNSR-----CPLNEEIVQARKIFPSVIKY 205  
 Db 158 DTEQERKRRLAKNRQOQIMNRFSSROOKFMDKHE-EYSAGND-----EDVDMOGEDLAG-- 1186

Db 167 DKEAW-----NHENCKGA-EDNGRLEDEDDHDKISKMLSVLIELFDHFDVFNQ 218  
 QY 206 VVE--MTWE-----BEKELPPELQIREK-----NER----- 230  
 Db 219 NIEPLTTIQKPLIAKRYFNPEREYEQADMLRLRAYRNQYMDDESSNKRHLTSDPLS 278  
 QY 231 ---YVCVLENDEHSHYVHSIORALDCELAQLHTTADKCGERRAVKAGA-YAACQ 285  
 Db 279 TLKDYAILVYDEFNYSOASAAIRQG-GPDKNKHIDLLTAKIDSEGRSLRCSADIASLM 337  
 QY 286 EAKEDIKSH--SENYSOHPLHVEVLHSEITMAHQFALRLGSSWMNKIMSY-SSDFRQIFCQ 342  
 Db 338 GRIFSQVSNGLSCTITQ---WYELHQEAC---KYSI---MWINDCLNPSTFQSLFRN 388  
 QY 343 A-----CLREEPD-----SENPCLI----- 357  
 Db 389 AIGKVLCSKYEPFYQSIDMTSVVRDYFSDSYLSDDPYLAHDHVSILGEGVKIPLGRHKSID 448  
 QY 358 -----SRL-----MLWDAKLYKGARKILHELIFSSFFMEMEYK 390  
 Db 449 PGDISAISPLNKNVIAEDHHEYNRLQYVLFLENRYWKKRKIVODLLIPTLASSAVOK 508  
 QY 391 KLFAFEFVKYKQLOKEYSDDHRSISITALS---VOMFTVPTLARHLIEE-----Q 440  
 Db 509 PMFTDQVLEIFPHMTRS--GTFMDREPOLTSRESVQLFTCTPTTAYSIFHSGHFNLIW 566  
 QY 441 NVISVITE--TLLE--VLPEYLDNRNKNFQGYSDOKLGRVYAVICDLKYI---LISKP 492  
 Db 567 SVIDVDFSTMDGTLVMQVRQSRNP--SKSYSISFKQGLYAVETLLSKITDNPILLKP 624  
 QY 493 TIWTERLMQFLEGFRSLKILT---CMOGMEI--RRQVGOHI-EVDPDWEAAIAIOMQL 547  
 Db 625 -----GEFIMIVTCLKLFNGAWKIKKREGEHVLRDQHFIPYLEVTTVS 669  
 QY 548 KNILLMF-----QEWCAACDELLVAYKECHKAVMRCSTSFSSSTVTVQSGCH----- 596  
 Db 670 YSIITFTKVLQOQSKDHIDORLLIGA-----INLLDSFL-----GHRNLSYK 711  
 QY 597 ---SLETSKYRVSEDLVSIHLPLSRTLAGL--HVRLS-RLGAVSRILHEVSEDFQFVVL 650  
 Db 712 LYKDFEIIKFKQISKEQVSFMNPVHTLFCFLVQHPVLSQVLSQKSYLVISDF----- 766  
 QY 651 VEYPLRCLVLVAQVVAEMWRNGLSLISQVFFYQDVKCREM--YDKDIIMLIQIGASLMD 708  
 Db 767 ---ALRSVVLCSQIDIGFWRNGMSVLIHOSAYI---KNPEMSSYSRDLQNLQ-AFLIE 819  
 QY 709 PNKFLLV---LQRYELAEAFNKTIQDQDLIKQYNTLIEEMQVLIIVIGERYVPGVG 765  
 Db 820 KNDFORVIYNMLDRWELLDWFDGSPSTETVYDDKISSIIQQFVAFLYQILVER----- 873  
 QY 766 NVTKREVTMRE-----IIHLICIEPMPSHATAKNLPEN-ENNETGLENVINKVATF 815  
 Db 874 DFYKKFDTLEETQLYNIKNAIIVKLYAEPLSVTDLLNDIPDLVLTESVSQFDTVLEEVSTY 933  
 QY 816 KPGVSGHGVYELKDESLED-----FNM--VFYHYSKTOHKSAAEMQKRRKQENKDEA 867  
 Db 934 IEPKSGKIWCFTQKEGTLOENRFLRLNMGNDFEHSATIVKSHLADSKKRAKIIYKPKQ- 992  
 QY 868 LPPPPPPPECFPAFSKVINLLNCDIMMYILRTVFERAIDTDSNLWTEGMQLQMAFHILALGL 927  
 Db 993 -----LLELDELDPCCAR---ELGSFTRTNLFA-----KLIFKLKULAV 1027  
 QY 928 LEEKQOLQKAPREEVTFDFYHKASRLGSSAMNQMILLEKLGIPQLEGOKDMITWILQMF 987  
 Db 1028 -----SDSSFSTYELLHLIHAIFRDEENVMGKDSLPEAVISKPICDLLLLSIV 1075  
 QY 988 DTVKRLREKSCULIVATTS-----GSESIKNDEITH-----DKE 1020  
 Db 1076 DSESGSFSEN--VWATADYLLDNMIMKRPATAVLESLETCFGTKYIADYKIRKANQGVNFE 1133  
 QY 1021 KAERKKAFAARLHRQKIMAOMSALOKNFIEHTKLMYDNTSEMPGKEDSIMESESIPAVS 1080  
 Db 1134 ETEQERKRRLAKNRQOQIMNRFSSROOKFMDKHE-EYSAGND-----EDVDMOGEDLAG-- 1186

```
QY 1081 DYSRIALGPKRGSVTEKEVLTCILCQEEQEVKIENNAMVLSACVQKSTALTQHRKPPIE 1140
Db 1187 -----ELNEFHGSLCHDD-----VSDDFFVPIPIYQNVSPVFLSNPTPME 1226
QY 1141 L-----SGEALDPLFMDPLAYTYTG-----SCGHVHAWCVQKQYF 1177
Db 1227 IYKPHGFDNNEHL--ATYNTDLFYKKKENGASQLMHSTQKVLVSCNHAHVYRCFKHYI 1284
QY 1178 EAVQLSSQRIHVDLFDLSEGEVCLPCKSLQNTVPIPIPLQPKQINSENADALQULTL 1237
Db 1285 D-----KARYSTDLF-----ICPLCQTCYNSVIPV-----DTVKLQSGDRLLQOKLT 1326
QY 1238 ARWIOVLARISGYNIRHAKGENPIPIFFNOGMDSTLPHSILSGVESSIKVYSIKE 1297
Db 1327 GGLDESLLTFEYSE-----CNDEVGKIL-----SLKDSNN--- 1360
QY 1298 MVILFATTIYRIGLKPDPDRPVPMLTWSTCAFTIOALE-----NLLDGEGRPLF 1349
Db 1361 -----GLRL--NRNDP-----TWIQDRFLTSLQFSNNICLLEMLSLNKNKDPF 1401
QY 1350 GALQNRQHNGLKALM-----OFAVAQRITCPQVLIQKH-----LVRLL 1387
Db 1402 GTLLSGEQKFTQLNILKSLAVYTRLT-----KHEMYSNSMKIFVAIYQVIRFP 1454
QY 1388 SVVLNPKSEDPT--PCLLSIDLPHVLGAVLAPPSLY--WDDPVDLQPSVSSYNHLYL 1443
Db 1455 RVETVLRSLRSLFKDCLQEV--LVERLKLGLTKDFGSGFYRKYESLRAQTCLDSSEFS--- 1509
QY 1444 PHLITMAHMLQILLTVDTLPLAQVOEDSEEAHSASSFFAEISQYTSGSGICDIPGWLW 1503
Db 1510 -----IVLKTILAGFG-----DQVKHTLDLFY----- 1534
QY 1504 VSLKNGITPLRCAALFFHYLLG-VTPPELHTNSAEGEYSALCSYLSLPTNLFLLFQY 1562
Db 1535 TELISELLPTLRSLIILLKALKQFMGTGGDDDFNEKDVLSGLASE-SREKHFHLL- 1589
QY 1563 WTVRPLRQRCADPALLNC-----LKGKNTVRYPRKNSLIELPD---DYSCLLNQASHF 1616
Db 1590 ---IRFLQOTDYD--LLMNSHSPLESSELVNAHDYCSIIKTDLATHLNTVYTNKNI 1645
QY 1617 RCPRSADDERKHVP-----LCLFCGAILCSNICQCEIVNGSEVGACIFHALHCK--- 1666
Db 1646 TLREENDQKIRTVNRLDYKICLIGCVKHAR-----TDGLEMOK---HMRCSHGSS 1695
QY 1667 -----ARGCAY-----PAPYLDXYGETD-PGLKRGNPPLHLRERYRKLHLVW 1707
Db 1696 GFLIPNISQVCLYLSRPDCTVNISAPYLSHSGESGRNAIERGDLTVLNHARYEHLTRLW 1755
QY 1708 QQHCLIEETAR 1718
Db 1756 ISNGIPGYISR 1766
```

## RESULT 5

```
GOG4_HUMAN
ID GOG4_HUMAN STANDARD; PRT; 2230 AA.
AC Q13439; Q13270; Q13654; Q14436;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Golgi autoantigen, golgin subfamily A 4 (Trans-Golgi p230) (256 kDa
golgin) (Golgin-245) (72.1 protein).
GN GOLGA4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX MEDLINE=96215236; PubMed=8626529;
RA Erlich R., Gleeson P.A., Campbell P., Dietzsch E., Toh B.-H.;
RT "Molecular characterization of trans-Golgi p230: a human peripheral
```

```
RT membrane protein encoded by a gene on chromosome 6p12-22 contains
RT extensive coiled-coil alpha-helical domains and a granin motif.";
RL J. Biol. Chem. 271:8328-8337(1996).
RN [2]
RP SEQUENCE FROM N.A.
RA Seelig H.P.;
RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 131-2230 FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=96125112; PubMed=8537393;
RA Fritzler M.J., Lung C.-C., Hamel J.C., Griffith K.J., Chan E.K.L.;
RT "Molecular characterization of golgin-245, a novel Golgi complex
protein containing a granin signature.";
RL J. Biol. Chem. 270:31262-31268(1995).
RN [4]
RP SEQUENCE OF 524-672 FROM N.A.
RC TISSUE=Gastric fundus;
RA Balague C.;
RL Thesis (1994), Instituto municipal de investigacion medica, Spain.
CC -!- FUNCTION: May play a role in vesicular transport from the trans-
Golgi.
CC -!- SUBCELLULAR LOCATION: ASSOCIATED WITH THE CYTOPLASMIC FACE OF THE
GOLGI MEMBRANE.
CC -!- ALTERNATIVE PRODUCTS:
Event=Alternative splicing; Named isoforms=4;
Comment=Additional isoforms seem to exist;
Name=1;
IsoId=Q13439-1; Sequence=Displayed;
Name=2;
IsoId=Q13439-2; Sequence=VSP_004272, VSP_004273;
Name=3;
IsoId=Q13439-3; Sequence=VSP_004274;
Name=4;
IsoId=Q13439-4; Sequence=VSP_004275;
-!- DISEASE: ANTIGEN IN THE AUTOIMMUNE DISEASE SJOREGREN'S SYNDROME AND
IN HEPATITIS B.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announcement/
or send an email to license@isb-sib.ch).
CC -----
EMBL; U41740; AAC50434.1; -
DR EMBL; X82834; CAA58041.1; -
DR EMBL; U31906; AAC51791.1; -
DR EMBL; X76942; CAA54261.1; -
DR Genew; HGNC:4427; GOLGA4.
DR MIM; 602509; -
DR GO; GO:0005802; C:Golgi trans-face; TAS.
DR GO; GO:0016192; P:vesicle-mediated transport; TAS.
DR InterPro; IPR000237; GRIP_domain.
DR Pfam; PF01465; GRIP; 1.
KW Golgi stack; Antigen; Coiled
FT DOMAIN 133 237
FT COILED COIL (POTENTIAL).
FT DOMAIN 276 1011
FT COILED COIL (POTENTIAL).
FT DOMAIN 1033 1214
FT COILED COIL (POTENTIAL).
FT DOMAIN 1259 2152
FT COILED COIL (POTENTIAL).
FT VARSPLIC 2154 2185
FT TPYKGNLYHTDVSLSFGPTEFEYLRKVLFEY -> HLTKV
FT AICTIRMSHSLNPLNSICEKCLSI (in isoform
FT 2).
FT /FTId=VSP_004272.
FT Missing (in isoform 2).
FT /FTId=VSP_004273.
FT Missing (in isoform 3).
FT /FTId=VSP_004274.
FT FTSPRSGLF -> SWLRSSS (in isoform 4).
FT /FTId=VSP_004275.
FT R -> K (IN REF. 3).
FT Y -> H (IN REF. 3).
```

FT	CONFLICT	276	276	T -> A (IN REF. 3).
FT	CONFLICT	584	584	K -> E (IN REF. 3).
FT	CONFLICT	628	628	T -> A (IN REF. 3).
FT	CONFLICT	630	630	K -> N (IN REF. 3).
FT	CONFLICT	682	682	K -> N (IN REF. 3).
SQ	SEQUENCE	2230	AA; 261139 MW; 3BB733DBIEA06134 CRC64;	

  

Query Match		2.3%;	Score 208;	DB 1;	Length 2230;
Best Local Similarity		18.1%;	pred. No. 0.00022;		
Matches		192;	Conservative 204;	Mismatches 403;	Indels 264; Gaps 38;
QY	188	NEEVIVQARKIFPSVIKYVVMETIWEKEKPELPELQIREKN-BRYXCVLVFNDDHHSYDVH	246		
Db	474	SEBQIAKQLKH-----EKELARKEQELTKLQTREREFQOMKVALEKQSEYSLKI	525		
QY	247	IYSLQALDCELAEQAOLHTTAIDKEGRRAVK-----AGAY-----AACQAEKEDIKS	293		
Db	526	SQKEQESALAELELQKKAILTESENKURDLOQRAETVTRILELESLEKSLQNK	585		
QY	294	HSNVSOH-----PLHVEVLHSEI--MAHQKFAIRLGSWMNKI-----MSYSSD	335		
Db	586	QSKDLAVLHAEKNKHKNKEITVAVEXHKTELESKHOQDAL-----WTEKLOVLKQXQTE	641		
QY	336	FRQIFQACUREPDSNPCLII-SRLMLWDAKLYKGARKLHLELIFSPFME-----	386		
Db	642	MEK-----LREKCEQEKETLLDKKEIFIQAHIETEMNEKTLEKLDVKQTELESSELSE	695		
QY	387	-----MEYKKLFAMEFYKYLQKQKEYISDDHDSISIT--ALS	423		
Db	696	VLKARHKEELSVLKQDQTKMQOELEAKMDEQKNHHQOQVDSIIKEHYSIORTEKALK	755		
QY	424	VOQFTVPTLA-----RHLEBQNVISVITETLLEVLPEYLDNRNKNF--FOGYS-----	470		
Db	756	DOINOLELLIKERDKHLKHEQAHVENLEADIKSEGELOQASAKLDVFOYSQSAHQETK	815		
QY	471	--ODKGRVYAVICDL--KYILISKPTIWTPLRMQF---LEGFRSFLKILTCMQGMEEI	523		
Db	816	AYEBQLAQLOQLDLLETERILLTKQVAEVAEQKDKVCTELDAHK---IQVDLMQOLEKQ	873		
QY	524	RRQVGQHIE-----VDPDWEAAATAIQMLKNILLMQEWACDEBELLVAYKECHKAVMR	578		
Db	874	NSTEMEQVKSLTVQYESKLEDGNKQEQDTKOILV-----EKENMILOMRGQKKEIE	925		
QY	579	CSTSFSSSKTVQSCGHSLETK-----	609		
Db	926	ILQTKLSAKEDSTHILNEEYETFKNQEKMEKVQKAKEMQETLKKLLDQEAJLAKEL	985		
QY	610	VSTHPLSRTLAGLHVRLSLRG-----AVSRLHEFVSFEDQVEVLVEYPLRCLVL	660		
Db	986	ENTALELSQEKQFNAKMLEMAQANSAGISDAVSRLR---TNQEQIESLETVHRREL--	1040		
QY	661	VAQVVAEMWRNGLSLISQVFFYQDVKYCREMDYDKIIMLOIGASLMDPNKFLLLVLQRY	720		
Db	1041	--NDVISIBKKLUNQQAEELOETHETQLQEK--EQEVAELK-----OKILLFGCEKE	1088		
QY	721	ELAEAFNKITISKDQDLIKQNTYLIEEMLOVLIIYIGVERVPGVGNVTKKEVTMRETIHL	780		
Db	1089	EM-----NKEITWLKEEGVKQDITLNELOEQ-----KQSAHVNSLAODETKLKAHLEK	1138		
QY	781	LCIEPMPHSAIAKNLPN-----ENNETHGLENVINKVATFKKPGVSGHGV	826		
Db	1139	LEVD-----LNKSLKENTFLQEQVLKMLAEDRKVKSELTSKLTDTDEEQSLKSSH	1192		
QY	827	ELKDESLKDNWFFYHYKTOHSAEHMQKRRKQENKDEALPPPPPPFCPAPSKVINL	886		
Db	1193	EKNKSLEDSLEFFKLSLEELAIQDITCCK-----KTEALLEAKTNELINISSSKTNA	1245		
QY	887	LNCDINMYILRTVFERAIDTDSNLWTBGMQLMAFHILALGLLEKQOLOKAPR-----	940		
Db	1246	ILSRISHCQHRT-----VKYEALLIKTCTVSELAQRLQTEQNTLNI	1290		
QY	941	---EVTPDFYHKASRLGSSAMNIQMLLEKLGIPQLEGQDKMITWILQMFDTVKRLREKS	997		

Dd    1291 SFOATHQLBEKENQKSMKADIESLVTKEALQEGGNQQAA-----SEKS 1333

Qy    998 CLIVATTSGESIKNDEITHDKKAERKRKAEARLRHQ-----KTMAQM 1042  
       | : || :| : :|| :| :| :| :| :| :| :| :| :| :| :| :|  
Db    1340 CITQLKKELSEN--NAVTLMKLELK-KVEISSLSKLTDNLVOLNSISLSREAAI 1396  
  
Qy    1043 SALQNKFIEHKLMYDNTSEMPCSKEDSIMEE-STRPAVSYS 1083  
       | :|| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|  
Db    1397 SSLRKQYDEEKCELLDQVDFSLFKVDFTLSKEKISALESQVDWS 1439  
  
RESULT 6  
GIAN\_HUMAN STANDARD; PRT; 3259 AA.  
ID GIAN\_HUMAN ID MEDLINE=9418728; PubMed=7511208;  
AC Q14789; O14398;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Golgi autoantigen, golgin subfamily B member 1 (Giantin) (Macrogolgtn)  
DE DE [Golgi complex-associated protein, 372-kDa] (GCP372).  
GN GOLGBL.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxId=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=9418728; PubMed=7511208;  
RA Seelig H.P., Schranz P., Schroeter H., Wiemann C., Griffiths G.,  
RA Renz M.;  
RT "Molecular genetic analyses of a 376-kilodalton Golgi complex membrane  
protein (giantin).";  
RL Mol. Cell. Biol. 14:2564-2576(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=94257116; PubMed=8198703;  
RA Seelig H.P., Schranz P., Schroeter H., Wiemann C., Griffiths G.,  
RA Renz M.;  
RT "Macroglolin -- a new 376 kD Golgi complex outer membrane protein as  
target of antibodies in patients with rheumatic diseases and HIV  
infections.";  
RL J. Autoimmun. 7:67-91(1994).  
RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=95100974; PubMed=7802676;  
RA Solda M., Misumi Y., Fujiwara T., Nishioka M., Ikebara Y.;  
RT Molecular cloning and sequence analysis of a human 372-kDa protein  
localized in the Golgi complex.;  
RL Biochem. Biophys. Res. Commun. 205:1399-1408(1994).  
CC -!- FUNCTION: May participate in forming intercisternal cross-bridges  
of the Golgi complex.  
CC -!- SUBUNIT: Homodimer; disulfide-linked.  
CC -!- SUBCELLULAR LOCATION: Membrane-associated protein. Golgi.  
CC -!- DISEASE: Antigen in chronic rheumatoid arthritis and in the  
autoimmune disease Sjogren's syndrome.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/or send an email to license@isb-sib.ch).

EMBL; X75304; CAAS3052.1; -.  
DR EMBL; D25542; BAA05025.1; -.  
PIR; A56539; A56539.  
DR PIR; I52300; I52300.  
GeneW; HGNC:4429; GOLGBL.  
DR MIW; 602500; -.  
DR GO; GO:0000139; C:Golgi membrane; TAS.  
DR GO; GO:0005795; C:Golgi stack; TAS.  
GO; GO:0016021; C:integral to membrane; TAS.

DR GO:0007030; P:Golgi organization and biogenesis; TAS.

KW Golgi stack; Antigen; Coiled coil; Transmembrane

FT DOMAIN 1 3235 CITOPLASMIC (POTENTIAL).

FT TRANSMEM 3236 3256 POTENTIAL.

FT DOMAIN 3257 3259 LUMENAL (POTENTIAL).

FT DOMAIN 48 593 COILED COIL (POTENTIAL).

FT DOMAIN 677 1028 COILED COIL (POTENTIAL).

FT DOMAIN 1062 1245 COILED COIL (POTENTIAL).

FT DOMAIN 1301 1779 COILED COIL (POTENTIAL).

FT DOMAIN 1828 3185 COILED COIL (POTENTIAL).

FT DOMAIN 2420 2423 POLY-GLU.

FT DOMAIN 2993 2996 POLY-SER.

FT CONFLICT 1 39 MISSING (IN REF. 3).

FT CONFLICT 215 215 A -&gt; AOLSSM (IN REF. 3).

FT CONFLICT 1765 1765 D -&gt; G (IN REF. 3).

FT CONFLICT 2950 2950 H -&gt; D (IN REF. 3).

SQ SEQUENCE 3259 AA; 376075 MW; 60376A20D8A178DD CRC64;

Query Match

Best Local Similarity 2.1%; Score 190.5; DB 1; Length 3259;

Matches 221; Conservative 207; Mismatches 424; Indels 377; Gaps 49;

QY 169 VNHEPRAGTIKENSRCPLNEEVIVQARKIFPSVIKYVVMETIWEBEKELPPELQ--IRE 226

Db 12 VLHLSGGDDTDQNMRAPLDPELHQSDMEFNNTQOEDVQERLAYAE-QLVVELKDIQ 70

QY 227 KNERVCYLFNDEHSHYDHVIVSYLQALDCELAELAQHMT-----TADK--EGRRVAKGA 280

Db 71 KDQVL-----QOKDALOEERKAADNKIKKLHAKAKLTSLNKYIEEMKAOGTV 121

QY 281 YAAOCEAKEDIKSHENVSOHPLHVEVLHSEIMAHOKFALRLGSWWNKINSYSDFRQIF 340

Db 122 LTFEQSEQLSKHDKSSTEEEMETIKKHLOKEELISTL----- 163

QY 341 COACLEEPSDENPLISLMLMDAKLYGARKILHELIFSPFMEYMKKLPAMFVKY 400

Db 164 -QAQL-TQAQAEPQASS-----TEME-----EFVVM 188

QY 401 YKQLOKEYISDDHDSRISITALSVMQNTVPTLARHLIEBQN-----VTSVIT 447

Db 189 KOOLQEK-----EFFISTLOALSQTQAQAQVVRKEDARFQTVRLHEDELQVLT 242

QY 448 ETLLEVLPEYLDNNKFNQGSQDKLGRVYAVICDKYLLIS----KPTIWERLRM-- 501

Db 243 QADVETEMQOKLURVLRKLEHEESLVGR--AQVDDLQOELTAABQRNQLSQOQQME 300

QY 502 -----QPLEGFRFLKILTCMQGMEIRROVGOHIEVDPDWEAAIAIQMLKNILLMF 554

Db 301 AEHNTLRNTVTEREESKILLEMELEVAERKLSFH-----NLQEEHHLLLEQF 349

QY 555 QEWCADEEL--LLVAYKECHKAVMRCSTSF1-SSSKT--VVQSCGHSLETYSRYSEDL 609

Db 350 EQAQQAQAELESRYSALEQKHKAEMBEKTSHLSLQKTGOELQSACDALQDQSKLLQDK 409

QY 610 VSIHPLSRTLAGLHVRLSRGA-----VSRL-----HEP---VSFDFQVEVLVEPLRC 657

Db 410 NQAVQSAQTIQOEDLOAQKSEISQFLNRLPLQOQHETASQTSFPDVPVNEGTVTEEN 469

QY 658 LVLVAQVVAEMRRNGLSLISVYFYVDVKCREMYDKDILMLQI-----GAS 705

Db 470 IASLQKRVVLENEKAGLLSSI-ELEELKAENEKLSQITLLEAQNRTEADREVSEIS 528

QY 705 LMD-PNK-----FLLVLVQRYELAE 725

Db 529 IVDIANKSSSAEESQDVLENTFSQKHKLSVILLMEKAEQEEIAFLKLOQKRAEEA 588

QY 726 FNKTISTKDDLIKQY--NTLIEMLQVLIYIVGERY----- 760

Db 589 DHEVL---DQEKWKQMEGEGAPIKMKVPLEDTGQDFPLMPNEESSLPAVEKQASTEHQ 645

QY 761 -----VPGVGNVTKVEVMNREIHLHLLCTEPMPSHA- 790

Db 646 SRTSEEISLNDAGVELKSTQDKGKSLSAVVDIGQCHQDELE-RLKSQLIELELNFHKAQ 704

QY 791 --IAKNLPENNETGLENVINKVATFKKPGVSGHGVY-----ELKDESL---KDFNMYP- 840  
Db 705 EYEKNLDEKAKEISNLNQLIEE---FKKNADNNSAFTALSEEDQLLSQVKELSMYTE 761  
QY 841 --YHYSTQHSKAEHMKRRKQENK---DEALPPPPPEFCPAFSKVINLLNCIDIMYI 895  
Db 762 LRAQVKOLEMNLAEARQRRLDYESQTAHDNLLTE-----QIHSLSEAKSKDVKEV 814  
QY 896 LRTVFE-----POLEGOKDMITWLOMFDVTKRLREKSLCLIVATTSSESTKNDIEIHD 921  
Db 815 LQNELDDVQLQFSEOSTLIRSLQSQONKESEVLEGAERVRHISSEKVEELSQAQSKLE 874  
QY 922 ILALG--LLBEK---QOLOKAPEE-----EVTDFYHKASRLG---SSAMNIQMLLEK 966  
Db 875 ITRKMDQLLEKKRDVETLOQTIEEKDQVQTEISFSMTKEMVQLNEEKFSLGVEIKLRBQ 934  
QY 967 LKGI-----POLEGOKDMITWLOMFDVTKRLREKSLCLIVATTSSESTKNDIEIHD 1018  
Db 935 LNLSSRAEAKKEQVEDNEVSSGLKQNYDEM-----SPAGQISKELQHE 980  
QY 1019 -----KEAERKRKAEEAARLHRQKIMQMSALQKNFIETHKLMDNTSEMPGKEDSIMBE 1073  
Db 981 FDLKLENEQKRKLQALINRKELLQVSRLEE---ELANLKDESKKEIPLSETERGEV 1037  
QY 1074 ESTPAVDYSRIALGPKRGPSVTEKEVLTCLQBE-----EQEVKIENNAMVLS- 1122  
Db 1038 EEDKENKEYSE-----KCVTSK-----COEIIYVLKQTISEKEVELOHIRKDLBE 1082  
QY 1123 --ACVOKSTALTOHRGKPIELSGEALDPL 1149  
Db 1083 KLAEEQFQALVKOMNOTLQDKTNQIDLL 1111

RESULT 7

SNEL\_HUMAN STANDARD; PRT: 8797 AA.  
AC Q9NF91; Q94890; Q8N9P7; Q8TCPI; Q8MMW6; Q8MMW7; Q8WFX6; Q96N17;  
AC Q9C0A7; Q9H525; Q9NS36; Q9NU50; Q9UJ06; Q9UJ07; Q9ULF8;  
DT 15-SEP-2003 (Rel. 42, Created)  
DT 15-SEP-2003 (Rel. 42, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Nesprin 1 (Nuclear envelope spectrin repeat protein 1) (Synaptic  
DE nuclear envelope protein 1) (Syn-1) (Myocyte nuclear envelope protein  
DE 1) (Myne-1) (Enaptin).  
GN SYNE1 OR MYNE1 OR KIAA0796 OR KIAA1756 OR KIAA1262.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP MUTAGENESIS OF 8758-LEU-CYS-8763.  
RC TISSUE=Heart, Placenta, Skeletal muscle, Spleen, and Testis;  
RX MEDLINE=21652858; PubMed=11792814;  
RA Zhang Q., Skepper J.N., Yang F., Davies J.D., Hegyi L., Roberts R.G.,  
RA Weissberg P.L., Ellis J.A., Shanahan C.M.;  
RT "Nesprins: a novel family of spectrin-repeat-containing proteins that  
RT localize to the nuclear membrane in multiple tissues.";  
RL J. Cell Sci. 114:4485-4498(2001).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORM 1), AND SUBCELLULAR LOCATION.  
RC TISSUE=Heart, Spleen, and Testis;  
RX MEDLINE=22296983; PubMed=12408964;  
RA Zhang Q., Ragnauth C., Greener M.J., Shanahan C.M., Roberts R.G.;  
RT "The nesprins are giant actin-binding proteins, orthologous to  
RL Drosophila melanogaster muscle protein MSP-300.";  
RN Genomics 80:473-481(2002).  
RN [3]  
RP SEQUENCE FROM N.A. (ISOFORM 4), AND VARIANT GLY-8323.  
RA Braune S., Abraham S., Padmakumar V., Tunngal B., Noegel A.A.,  
RA Korenbaum E.;  
RT "The longest isoform of enaptin/Syne-1, a nuclear envelope associated

RT protein, binds actin cytoskeleton via the alpha-actinin-like actin-binding domain.";

RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.

RN [4]

RP SEQUENCE FROM N.A. (ISOFORM 8 AND 9).

RA Zhang Q., Shanahan C.M.;

RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.

RN [5]

RP SEQUENCE FROM N.A.

RA Almeida J., Clark S., Griffiths C., Lloyd D., Parker A., Smith M., Tracey A., Williams S.;

RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.

RN [6]

RP SEQUENCE OF 1-856 FROM N.A.

RC TISSUE=Kidney;

RA Gough L., Fan J., Lisa G., Chu S., Winnick S., Beck K.A.;

RL "Golgi localization of syne-1.";

RT Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.

RN [7]

RP SEQUENCE OF 28-778 AND 2901-3476 FROM N.A.

RC TISSUE=Adrenal gland, and Teratocarcinoma;

RA Ninomiya K., Wagatsuma M., Kikkawa E., Omura Y., Yokoi T., Kodaira H., Furuya T., Takahashi M., Kikkawa E., Omura Y., Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M., Yamazaki M., Sugiyama T., Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B., Takemoto M., Ota T., Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.;

RL "NEDO human cDNA sequencing project.";

RT Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.

RN [8]

RP SEQUENCE OF 443-8797 FROM N.A. (ISOFORM 5).

RC TISSUE=Brain;

RA Nagase T., Kikuno R., Hattori A., Kondo Y., Okumura K., Ohara O.;

RL "Prediction of the coding sequences of unidentified human genes. XIX. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro.";

RT DNA Res. 7:347-355(2000).

RN [9]

RP SEQUENCE OF 743-8797 FROM N.A. (ISOFORM 6).

RC TISSUE=Brain;

RA Ansong W., Wirkner U., Mewes H.-W., Well B., Wiemann S.;

RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.

RN [10]

RP SEQUENCE OF 4219-8797 FROM N.A. (ISOFORM 7).

RC TISSUE=Brain;

RA Nagase T., Ishikawa K.-I., Kikuno R., Hirose M., Nomura N., Ohara O.;

RL "Prediction of the coding sequences of unidentified human genes. XV. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro.";

RT DNA Res. 6:337-345(1999).

RN [11]

RP SEQUENCE OF 6922-8797 FROM N.A.

RC TISSUE=Brain;

RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;

RL "Prediction of the coding sequences of unidentified human genes. XI. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro.";

RT DNA Res. 5:277-286(1998).

RN [12]

RP REVISIONS.

RC TISSUE=Brain;

RA MEDLINE=22158633; PubMed=12168954;

RA Nakajima D., Okazaki N., Yamakawa H., Kikuno R., Ohara O., Nagase T.;

RT "Construction of expression-ready cDNA clones for KIAA genes: manual curation of 330 KIAA cDNA clones.";

DNA Res. 9:99-106(2002).

RL [13]

RN SEQUENCE OF 7631-8797 FROM N.A., AND CHARACTERIZATION.

RA MEDLINE=21659781; PubMed=11801724;

RA Mislow J.M.K., Kim M.S., Davis D.B., McNally E.M.;

RT "Myne-1, a spectrin repeat transmembrane protein of the myocyte inner nuclear membrane, interacts with lamin A/C.";

RL J. Cell Sci. 115:61-70(2002).

RN [14]

RP SEQUENCE OF 8406-8797 FROM N.A.

RA Ma F.-R., Zhu L.-P.;

RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.

CC -! FUNCTION: Involved in the maintenance of nuclear organization and structural integrity. Probable anchoring protein which tethers the nucleus to the cytoskeleton. Connects nuclei to the cytoskeleton by interacting with the nuclear envelope and with F-actin in the cytoplasm.

CC -! SUBUNIT: Interacts with MUSK, with F-actin via its N-terminal domain, and with LMNA in vitro (By similarity).

CC -! SUBCELLULAR LOCATION: Type IV membrane protein (Potential). The largest part of the protein is cytoplasmic, while its C-terminal part is associated with the nuclear envelope, most probably the outer nuclear membrane. In skeletal and smooth muscles, a significant amount is found in the sarcomeres.

CC -! ALTERNATIVE PRODUCTS:

CC Event-Alternative splicing: Named isoforms-9;

CC Name=1;

CC IsoId=Q8NF91-1; Sequence=Displayed;

CC Name=2; Synonyms=Beta;

CC IsoId=Q8NF91-2; Sequence=VSP\_007130;

CC Name=3; Synonyms=Alpha;

CC IsoId=Q8NF91-3; Sequence=VSP\_007132, VSP\_007144;

CC Name=4;

CC IsoId=Q8NF91-4; Sequence=VSP\_007134, VSP\_007139, VSP\_007140, VSP\_007144;

CC Name=5;

CC IsoId=Q8NF91-5; Sequence=VSP\_007135, VSP\_007136;

CC Note=No experimental confirmation available;

CC Name=6;

CC IsoId=Q8NF91-6; Sequence=VSP\_007137, VSP\_007138;

CC Note=No experimental confirmation available;

CC Name=7;

CC IsoId=Q8NF91-7; Sequence=VSP\_007141, VSP\_007142;

CC Note=No experimental confirmation available;

CC Name=8; Synonyms=Beta 2;

CC IsoId=Q8NF91-8; Sequence=VSP\_007131;

CC Name=9; Synonyms=Alpha 2;

CC IsoId=Q8NF91-9; Sequence=VSP\_007133, VSP\_007143, VSP\_007144;

CC -! TISSUE SPECIFICITY: Widely expressed. Highly expressed in skeletal and smooth muscles, heart, spleen, and peripheral blood leukocytes.

CC -! DOMAIN: The Klarsicht domain, which contains a transmembrane domain, mediates the nuclear envelope targeting.

CC -! SIMILARITY: Belongs to the Nesprin family.

CC -! SIMILARITY: Contains 1 actin-binding domain.

CC -! SIMILARITY: Contains 2 calponin-homology (CH) domains.

CC -! SIMILARITY: Contains 12 HAT repeats.

CC -! SIMILARITY: Contains 1 Klarsicht domain.

CC -! SIMILARITY: Contains 31 spectrin repeats.

CC -! CAUTION: Ref.5 (CAB55865, CAB55866, CAC16280 and CAC16281) sequences differ from that shown due to erroneous gene model prediction.

CC -! CAUTION: Ref.7 (BAB71097) sequence differs from that shown due to a chimeric cDNA.

CC -! CAUTION: Ref.14 sequence differs from that shown due to two frameshifts in positions 8412 and 8784.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

Qy	915	----	-----MLQWAFHLLA-----	-----LGLLEK-----QQDQKAAPEEVEVDFYHKA	950	
Db	6033	CEADPA	BQALQSTLTVLAERMSTIRK	MAKSGKQQLLEKUNDQLEQ	QEQALQRYCEA	6092
Qy	951	SRUGSSAMNTMLLEKLGIP	QLEGQDMITWLOMPD	TVKLRKESCLIVATTS	SGSESI	1010
Db	6093	DEDSWLLSTKATIDTALSP	-----KEPMDMAQLMDCQNM	LVIEQKVVALS-----ELSV	6145	
Qy	1011	KNDIEI-----	THDKAEARKKAEAA	RLHRQKINQAQMSALQK	NFIETHKLMIYNTSEMPG	1065
Db	6146	HNENLLLEGKAHTKDEAEQL	-----ACKLRR-----	LKGSILLELQRALHDKQLNMQ	6192	
Qy	1066	-----KEDSIMEEST-----	PAVSDY-----	SRIALGPKRGPSTVEKVL	FCILQCEQVKIEN	1116
Db	6193	TAQEKESDYDLTATQSPG	VOEWLAQARTTWTQOR	SSLQQQKEL-----EQL- AEQ	6244	

D6	6245	K5LLRVSARGEIILQH--SAAETSDAGEKPDVL5QELGMEGKSSAEDQMR-MKWE5	6301
QY	1111/	NAWLSACQVSTAL2QHRGRPIELSEALD-PLFMDPDLAYGTYTSCGHVYHAVCWQK	1177

QY	1176	YFBAVLSSQORHVDLFDESEGYLCPCKSLCNTVPIPLQPKINSINADALAOQL	1233
Db	6302	LHQ--EFSTKQKLLQNVLQEQQEQL-----YSRPNLLSGVPLVKGDVPTQDKSAVTSLL	6355
QY	1236	T-LARMIQTVLARISGYNIRHAKGENPIPIFFNOGMGDSLTFPHSLISFGVSESIKYSNS	1294
Db	6356	DGNQAFEEVSSGG-----AKRQS---IHLFQKLYD-----GVSATSTWLDD	6396
QY	1295	IKEMWILFATYIRIGLKVPPDERDPRVPLMTWSTCAFTQATENLLG-----DEGKPL	1348
Db	6397	VEER--LFVAT-----ALLPETE-----TCLFQGEILAKDIKEMSEMDKNKNL	6439
QY	1349	FG-ALONRQHN-----GLKALMOFAVAQRITCPQVILQIKHLVRLLS	1388
Db	6440	FSQAFFPENGNDRVIEDTLGCLLGRLSLLDSVVNQNR--CHQ--MKERLQOOILN	6488

```

RESULT 8
CEP2_HUMAN
ID CEP2_HUMAN STANDARD; PRT; 2442 AA.
AC Q9BV73; O14812; O60588; Q9H450;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Centrosomal protein 2 (Centrosomal Nek2-associated protein 1) (C-NAP1)
DE (Centrosome protein 250) (Centrosome associated protein CEP250).
GN CEP2 OR CNAF1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A. (ISOFORM 1), AND AUTOIMMUNE DISEASE.
RP
RC TISSUE=Cervical carcinoma;
RC MEDLINE=98165428; PubMed=9506584;
RA Mack G.J., Rees J., Sandblom O., Balczon R., Fritzler M.J.,
RA Rattner J.B.
RT "Autoantibodies to a group of centrosomal proteins in human autoimmune
RL sera reactive with the centrosome.";
RL Arthritis Rheum. 41:551-558(1998).
[2]
SEQUENCE FROM N.A. (ISOFORMS 1 AND 2), INTERACTION WITH NEK2,
RP AND SUBCELLULAR LOCATION DURING THE CELL CYCLE.
RC TISSUE=Placenta;
RC PubMed=9647649;
RX
RY Fy C.A.M., Mayor T., Meraldi P., Stierhof Y.-D., Tanaka K., Nigg E.A.;
RT "C-Nap1, a novel centrosomal coiled-coil protein and candidate
RL substrate of the cell cycle-regulated protein kinase Nek2.";
RL J. Cell Biol. 141:1563-1574(1998).
[3]
SEQUENCE FROM N.A.
RX
RY MEDLINE=21638749; PubMed=11780052;

```







RA Dufosse-Laurent V., Ferron C., Lechaplais C., Louesse C., Muselet D.,  
RA Magdelenat G., Pateau E., Petit E., Sirvain-Trukniewicz P., Trybou A.,  
RA Vega-Czarny N., Bataille E., Bluet E., Bordelais I., Dubois M.,  
RA Dumont C., Guerin T., Haffray S., Hammadi R., Muanga J., Pellouin V.,  
RA Robert D., Wunderle E., Gauguet G., Roy A., Sainte-Marthe L.,  
RA Verdier J., Verdier-Discala C., Hillier L., Fulton L., McPherson J.,  
RA Matsuda F., Wilson K., Scarpelli C., Gyapay G., Winkler P., Saurin W.,  
RA Quetier F., Waterston R., Hood L., Weissbach J.,  
RA "The DNA sequence and analysis of human chromosome 14.";  
RA Nature 421:601-607(2003).  
RA [6]  
RA SEQUENCE FROM N.A. (ISOFORM 8).  
RA TISSUE=Brain;  
RA MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Heiton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalls D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,  
RA "Generation and initial analysis of more than 15,000 full-length human  
RA and mouse cDNA sequences.";  
RA Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RA [7]  
RA SEQUENCE FROM N.A. (ISOFORM 3), AND SEQUENCE OF 1-956 AND 5133-6885  
RA FROM N.A.  
RA TISSUE=Spleen, and Tongue;  
RA Jikuya H., Takano J., Nomura N., Kikuno R., Nagase T., Ohara O.,  
RA Ninomiya K., Wagatsuma M., Kanda K., Kondo H., Yokoi T., Kodaira H.,  
RA Furuya T., Takahashi M., Kikkawa E., Omura Y., Abe K., Kamihara K.,  
RA Katsuta N., Sato K., Tanikawa M., Yamazaki M., Sugiyama T., Irie R.,  
RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,  
RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,  
RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Murakawa K.,  
RA Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,  
RA Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.,  
RA "NEDO human cDNA sequencing project.";  
RA Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.  
RA [8]  
RA SEQUENCE OF 5709-6885 FROM N.A. (ISOFORM 2).  
RA TISSUE=Brain;  
RA MEDLINE=99246063; PubMed=10231032;  
RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Hirose M.,  
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;  
RA "Prediction of the coding sequences of unidentified human genes. XIII.  
RA The complete sequences of 100 new cDNA clones from brain which code  
RA for large proteins in vitro.";  
RA DNA Res. 6:63-70(1999).  
RA [9]  
RA REVISIONS.  
RA TISSUE=Brain;  
RA MEDLINE=22158633; PubMed=12168954;  
RA Nakajima D., Okazaki N., Yamakawa H., Kikuno R., Ohara O., Nagase T.;  
RA "Construction of expression-ready cDNA clones for KIAA genes: manual  
RA curation of 330 KIAA cDNA clones.";  
RA DNA Res. 9:99-106(2002).  
RA [10]  
RA SEQUENCE OF 5754-6885 FROM N.A.  
RA MEDLINE=21154917; PubMed=11230166;  
RA Wiemann S., Weil B., Wellenreuther R., Gassenhuber J., Glassl S.,  
RA Ansorge W., Boecker M., Bloecher H., Bauersachs S., Blum H.,  
RA Lauber J., Duesterhoeft A., Beyer A., Koehrer K., Strack N.,

Query Match 1.9%; Score 177; DB 1; Length 6885;

Best Local Similarity 18.2%; Pred. No. 0.094;

Matches	266;	Conservative	249;	Mismatches	505;	Indels	442;	Gaps	75;
Qy	174	GRA-GTIKENSRCPLNEEVIQVARKIPFVSIKYVEMTIWEEKELPELQIREKNERY	232						
Db	970	GRTKGLIKEHEAC-FSEEGCLYQLNHHEVRELCE-----ELP-----SKSQOEV	1015						
Qy	233	CVLFNDBHHYDVHYISYLRALDCELAEALHTTAIDKEGRRAVKAGAYAACQE-	286						
Db	1016	KRLKLD---YEQ---KIERLLKC---ASEIHTMLQPTAGTSKNEGTITTSENRGDGH	1065						
Qy	287	-----AKEDIKSHSENVSQPHLHVEV-----LHSEITMAHQKFAIRLGSMWNKIMSY-	332						
Db	1066	SEAPFAKSDNQPSTEKAMEPTMFKFSLASVLRPLQEEIMEKDYSAINSLLERYDTRDI	1125						
Qy	333	-----SSDRQIFCQACLRPEPSENPCLLSRLMLWDAKLYKGARKTLHELIF	380						
Db	1126	LEHLLQNNKFRITSDF-----SSEDRSSSCLOAKLT--DLQVIKNET-----	1166						
Qy	381	SSFFMEMEYKKLFAMEFVKYKQLOKEYISDDHD-----RSISITALSVMQFTVPTLAR	434						
Db	1167	DARKKEP---IISLKLLENHVNDIKKPFVIKERDTLKERERELQMT--LNTFMESLEFALR	1222						
Qy	435	HLIEEQNVISVITETLEVLPEYLDNRNNKFNFGYSQDKLGRVYAVICDLKYLISKPTI	494						
Db	1223	LVLPEKASLLCGSDLPL-----HKMAIQGFHLIDADRIYQHLRNIQ-DSIAQKIE	1273						
Qy	495	WTERLR-----WQFLEGF-----RSFLKILCMQGMEEIRQVQGHIEVDDPWEAIIA	542						
Db	1274	ICNRLEPGEFVJLKEHPFDLHAMQNIILKYTKQFEGMNRHVORSEDTLKALEDFLASLR	1333						
Qy	543	-----IQOMKILNLLMFQEWACDBELLVAVYKECH--KAVMRCSTSF	584						
Db	1334	TAKLSAPVTDLSASDTQVAQENTLVTKNK-----EGEIHLMKDKAKHLDRCLKMLDMSFK	1389						
Qy	585	SSSKTVVQSCGHSI-----ETKSYRVSDELVSIHPLSRTLAGELVHLSRLGAVSRL	636						
Db	1390	DAERGGDTSCENLDDAFSIKSETHGVGVQEEETENKLEACIFKKNELLKNIODVQSQ	1449						
Qy	637	HEFVSPEDFQEVVLVYPLRCLVLVLAQVAAEM--WRNGISLSIQVFIYQD--VKCREEM	692						
Db	1450	ISKIGLKDPVTPA--VKHRKSLTRLDKVLDEYBEEKRHLGEMANSLPHFDGKREKTVNQ	1508						
Qy	693	YDKDIIMLOIGASL---MDPNKFLLLVQRYELAEAFNKITSTKDODLIK-QYNTLIEE	747						
Db	1509	QNTVVLWENTKALVTCELCQGRVLELLKQYQNFKSIITLTLAQESVLSLAQSYMGKE	1568						
Qy	748	MLQVLIYIGERYVPGVGNVTKEBVTMREIHLILCIEPMPSHAIAKNLPENNETGTLEN	807						
Db	1569	NLAKRTAEI-----EIVKEEFNE---HLEVVDKI--NQVCKN-----LQF	1603						
Qy	808	VINKVATFKPGVSGHGVYELKDESLKDFNMVYFYHSKTOHSAEHMQKRRKQENKDEA	867						
Db	1604	YLNKMTFEEPPPEE-----KEANIIVDRW-----LDINEKTDYIENLGRA	1644						
Qy	868	LPPIPPPEFCPASKVINLNCNDIMMYILKTVTERAIDTDSNLWTEGMLO-MAPHILALG	926						
Db	1645	L-----ALWDKLFNLKN-----VIDE-----WTEKALQKLMELH---Q	1673						
Qy	927	LLEE---KQOLQKAPEEBVTDFYHKASRLGSSAMNIOMLEKLGIPQLSGQKDMIT	981						
Db	1674	LTEEDRERLKEELQ--VHEQKTSFESRRVA-----EIQFLQSSSE-IP-LELQV-MES	1721						
Qy	982	WILQMDTVKRLREKSLVATTSSGSEIKNDEITHDKAEKRKAARL-----HR	1035						
Db	1722	SILNKHEHVOK-----CLTGESNCHALSGSTAELRDQAKTOIGMTESLLKALSPDS	1776						
Qy	1036	QKIMQASALQKNFIETH-----KLMYDNTSEMPCKEDSIMEES	1075						
Db	1777	LEIFTKLEETQQQILQOKHSMILLENQIGCLTPELSFKQYESVDLFWTKSVLQDHF	1836						
Qy	1076	TPAVSD-----YSRIALGPGRK--GPSVTEKEV-----LTCILCQEEQEVKIENNAM	1119						

Db	1837	SKLLNDQCKNFNDWFSIKVNLKCFESETTKKSEVQKLOKLSDFLLEGRNSKIKQVDS	1899
Qy	1120	VLSACVQKSTALTQHRGKPIELSGEALDPLFMDPLAYGVTGSCGHVMAVCWQKYFEA	1179
Db	1897	VLK-----HVKKHLPK-----AHVKELISW-----	1916
Qy	1180	VQLSSQORIHVDLDESGEYLCPLCKSLCNCNTVPIIPLOPKINSENADALAQLL	1235
Db	1917	--LVGQE-----FELEKWSIC-----QARAKELEDSLQQLLRQD	1950
Qy	1236	---TLARQIOTVLARISYNTIRHAKGNPIPIFNQGMGSTLEFHSILSGVSESSIK-Y	1291
Db	1951	DHRNLRKWLITNQEEKWKGMEEPEGKTE-----LFCQALARKREQFESVAQ--LNNSLKEY	2003
Qy	1292	SNSIKEMVILLPATTI--YRIGLK--VPPDERDPRV-----MLTWSTCAFTI	1334
Db	2004	GFTEEEELIMATCIMDRYOTLLQQLSEIEEEDKLLETPEDOSFNDLAHDVIHW-----I	2057
Qy	1335	QAIENLL-----GDEGK-PLFGALQNROHNGIKALMQFAVAQRITCPOVLQKHLVRLLSV	1389
Db	2058	KEIKESLWLVNSSSGKMPLEERIQ-----IKEI-----ILLKPSGDARIET	2099
Qy	1390	VLPNIKESDPCLLSIDLFLHVLGAVLAIFSLYWDPDVLDQPSVSSSYNHLXFLHITM	1449
Db	2100	IMKQAESSEAP-----LVQKLTDISNQWDNTLHL-----ASTY-----L	2134
Qy	1450	AHMLQIILLVDVTGLPLAQVQED	1471
Db	2135	SHQEKLLL---EGEKYLOSKED	2153

RESULT 10

MY5A\_CHICK

MY5A\_CHICK

STANDARD;

PRT; 1829 AA.

AC Q02440;

DT 01-JUN-1994 (Rel. 29, Created)

DT 01-JUN-1994 (Rel. 29, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Myosin Va (Myosin 5A) (Dilute myosin heavy chain, non-muscle) (Myosin heavy chain P190) (Myosin-V).

GN MY5A.

OS Gallus gallus (Chicken).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.

OC NCBI\_TaxID=9031;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Brain;

RX MEDLINE=33012002; PubMed=1383040;

RA Sanders G., Lichte B., Meyer H.E., Killmann M.W.;

RA "cDNA encoding the chicken ortholog of the mouse dilute gene product. Sequence comparison reveals a myosin I subfamily with conserved C-terminal domains.";

RT FEBS Lett. 311:295-298(1992).

RN [2]

RP SEQUENCE FROM N.A., AND SEQUENCE OF 1154-1163.

RC TISSUE=Brain;

RX MEDLINE=93107155; PubMed=1469047;

RA Esprefacio E.M., Cheney R.E., Matteoli M., Nasclmento A.A., de Camilli P.V., Larson R.E., Mooseker M.S.;

RA "Primary structure and cellular localization of chicken brain myosin-V (p190), an unconventional myosin with calmodulin light chains.";

RT J. Cell Biol. 119:1541-1557(1992).

CC -!- FUNCTION: PROGRESSIVE ACTIN-BASED MOTOR THAT CAN MOVE IN LARGE STEPS APPROXIMATING THE 36-NM PSEUDO-REPEAT OF THE ACTIN FILAMENT. MAY BE INVOLVED IN MELANOSOME TRANSPORT, OR ALTERNATIVELY, IT MAY BE REQUIRED FOR SOME POLARIZATION PROCESS INVOLVED IN DENDRITE FORMATION (BY SIMILARITY).

CC -!- SUBUNIT: MAY BE A HOMODIMER, WHICH ASSOCIATES WITH MULTIPLE CALMODULIN OR MYOSIN LIGHT CHAINS.

CC -!- SUBCELLULAR LOCATION: GOLGI-DERIVED CYTOPLASMIC MEMBRANES

CC CC (POTENTIAL).

CC CC -!- TISSUE SPECIFICITY: NEURONAL AND NONNEURONAL CELLS OF THE BRAIN.

CC CC -!- SIMILARITY: Contains 1 myosin-like globular head domain.

CC CC -!- SIMILARITY: Contains 1 dilute domain.

CC CC -!- SIMILARITY: Contains 6 IQ domains.

CC CC -----

CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC CC the European Bioinformatics Institute. There are no restrictions on its

CC CC use by non-profit institutions as long as its content is in no way

CC CC modified and this statement is not removed. Usage by and for commercial

CC CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC CC -----

DR EMBL; X67251; CAA47673.1; -

DR EMBL; Z11718; CAA77782.1; -

DR PIR; S19188; S19188.

DR HSSP; P10587; 1BR2.

DR InterPro; IPR002710; DIL.

DR InterPro; IPR000048; IQ\_region.

DR InterPro; IPR001609; myosin\_head.

DR Pfam; PF01843; DIL; 1.

DR Pfam; PF00612; IQ; 6.

DR Pfam; PF00663; myosin\_head; 1.

DR PRINTS; PR00193; MYOSINHEAVY.

DR ProDom; PD003376; DIL; 1.

DR ProDom; PD000355; myosin\_head; 1.

DR SMART; SM00015; IQ; 6.

DR SMART; SM00242; MYSC; 1.

DR PROSITE; PS50096; IQ; 6.

DR Myosin; Repeat; ATP-binding; Calmodulin-binding; Actin-binding;

CC CC Coiled coil; Phosphorylation.

CC CC MYOSIN HEAD-LIKE.

FT DOMAIN 1 766

FT DOMAIN 767 789

FT DOMAIN 790 814

FT DOMAIN 815 837

FT DOMAIN 838 862

FT DOMAIN 863 887

FT DOMAIN 888 915

FT DOMAIN 916 1239

FT DOMAIN 1315 1419

FT DOMAIN 1661 1766

FT NP\_BIND 163 170

FT DOMAIN 644 666

FT MOD\_RES 1734

FT CONFLICT 1142 1142

CC CC E -> EQ (IN REF. 2).

CC CC SEQUENCE 1829 AA; 212381 MW; 0538B278DFC09F6E CRC64;

Query Match 1.9%; Score 176; DB 1; Length 1829;

Best Local Similarity 18.0%; Pred. No. 0.015;

Matches 283; Conservative 248; Mismatches 578; Indels 464; Gaps 76;

Qy 99 GAFOLCGRVFKSGTETYSRCD---CAI-----DPTCVLC---MDCFDQSVHKNHRYKMH 147

Db 324 GIFRILAGILHLGNVFRASRSDSCAIPPKHDLPTIFCDLMGVD-YEEMAWLCHRKLAT 382

Qy 148 STGGGFCDCGTEAWKTPFCVFNHEPGRAGTIKENSRCPLNEEVIVQARKIPPSVIK--- 204

Db 383 A-----TETY-IKPISKHAINARDALAKHIYANFNWIVDHNKALHSTVKQHS 431

Qy 205 YVEMTWEEKELPPELQIREKNERYCYVLNDE--HHSYDHYIYSLQRALDCLAEQA 262

Db 432 FIGVDIYGF-----TPEINSFEQ--FCINYANEKLOQFNMHVFKLEQE---EYMREQ 481

Qy 263 LHTTAIDKEGRR-----AVKAGAYAAOEAKEKIDKSHSENVSHPLHVEVLHSEIMAHQ 316

Db 482 IPWTLDIFYDNPICINLEAKMGVLDDLDECKWPKGSDDTWAQKLYNTHL-----N 533

Qy 317 KFAL-----RLGSWMNK--IMSYSSDFRQIFQACLRPEEDSENPCILSRMLWDKLYKG 370

Db 534 KCALFEKPRLS--NKAFIINKHFAKVEYQCEGFLEKNKD-----VYEE 575

Qy 371 ARKILHELIFSFTMEMEYKKLFAMEFVYKYLQLOKEIYSDHDSRISITALS-----VQM 426

Db 576 QIKVL-----KSSKKFKLLPE--LFODEKAISPTSATPSGRVPL 613

Qy 427 FTVF-----TLARHL-----IEQNVISVITETLLEVPDEYLD--RNNKFNFGYS 470

Db 614 SRTVPKAKARPGQTSKEHKKTGVGHFRNSLHLLMETLNATTPHYVRCIKPNDPKFPPTF 673

Qy 471 QDKLG-----RVYAVICDLAYILISKPTIWERLRMOFLGFRSFLK---IL-----TCMQ 518

Db 674 DEKRAVQQLRACGVLETIRISAAGFSPRWYQ---EFFSKRYVLMKOKDVLSDRKOTCKN 730

Qy 519 GMEEI-----RRQVG-----OHTIEVDPDWEAAIAIOMQLKNILLMFOWCA 559

Db 731 VLEKLILDKDKYQFGKTKIFFRAGQVAYLEKTRADKLRAACIRIQTRGWLARKYMRM 790

Qy 560 CDEELLVAYKECHKAMVRCSTSFSSKT--VVQSCGHS--LFTKSYRVSIEDLVSHLPL 616

Db 791 RRAAITIQRVVRGHOA--RCYATFLRTRAAIIIQKFORVMYVVRKRYQCMRDAT---IAL 845

Qy 617 SRTLAGLHVRLSRLGAVSRLEHFVSFEDFQVEVLVEYPLRCLVLVAQVVAEMRRNGLSL 676

Db 846 QALLRGYLVN--NKYQMLREHKS-----IIQKHVRGW-----877

Qy 677 ISOVEFYQDVKC-----REEMYDKDIIMLIQIGASLMDPNKFLLLVLQRYELAEAFNK 728

Db 878 LARVHYHRTLKAIVILQCCYRRMMAKRELKLUKLEARSERYKKLHIGLE-----NK 929

Qy 729 --TISTKQODLIKQYNTLIEEM--LOVLIYIVGERYPVG-----GNVTKEEVTM 774

Db 930 INQLQKIDEQNKYKSLLEKNLNLEITYSTETEKLRSDVERLRMSEEAKNATNRVLSL 989

Qy 775 REIIHLCLIEPMPHSAIAKNLPENENNETGLENVINKVATFKPGVSGHGVYELKDESL- 833

Db 990 QEEI-----AKRLKELHQTOQTEKKTIEWADKYHETEQLVN-----ELKQNTL 1034

Qy 834 -----KDFNMVYHYSKTQHSKAEHMKK---RRKQ---ENKDEALPPPPPPPPCFPAFSK 882

Db 1035 LKTEKEELNRRHDOAK---EITETWEKKLVETKOLELDLNDERL-----R 1078

Qy 883 VINLLN-----CDIMMYILRTVPERAIDTDSNLWTEGMLQMAFHILA---924

Db 1079 YQNLLNEFSRLEERYDDLDKDENNLMVSIKPGHKRTDSTHSS--NESEVTFSSSEITAEAD 1136

Qy 925 IGLLEEKQOLQKAPEEVTFDFYHKASRLGSSAMNOMLLEKLGIPQLEGQKDMTWIL 984

Db 1137 LPLRMEEPSEKKAP-----LDMSLFLKQKRYTELEQEKOSLQ---1174

Qy 985 QMFDTVKRLREKSLCLIVATTSGSEITKNDIETHDKAEKRAEAAARLHROKIMQMSA 1044

Db 1175 ---DELDRKEEQALRAKAKEERPPIRGAEEYESLK---ROELES---ENKKLKNELE 1225

Qy 1045 LQKNFIETHKLMYDNTSEMPGKEDSTIMEEESTPAVDYSRIALGPKRGPVTEKEVITCI 1104

Db 1226 LQKALTETR-----APEVTAQAPAY-RVLL-----DQLTSV 1256

Qy 1105 LCOEEOEVKIENNAWVLSACVOKSTALTQHRGKPIELSGALDPLMDPDLATGYTGSC 1164

Db 1257 --SEELEVEKEEVLIRLSOLVSQKEAI-----QPKEDKNTMTDSTILLEDVQKMKOGEI 1309

Qy 1165 GHVMAVCQKYFEAVQLSSQORIHVDFLDESSEYLCPLCKSL-----CNTVPIPI 1217

Db 1310 AQAYIGLKTNRLLSESQOLSKSHENELESRGE-----IQSLKEENNQQOQLLANLQ 1364

Qy 1218 LQPO-----KINSNADALAQL-----LTLARMTQTVLARIISGYNIRHAKG 1258

Db 1365 LPEAREIASLOHEITRLTNENLDLMEOLEKQDKTVRKLLKQKLVAKKILEVGOEVEN 1424

Qy 1259 ENPIPIFFNOGMDSTL-----EFHSILSFGVESIK-----1290

Db 1425 ISPGQII-----DEPIRPVNIPRKEKDFQGMLEYKKEDEQKLVKNLLELKPGRVAVNL 1478

Qy 1291 -----YNSNKEWVILFATTIYIRGLKVPVPPDERPRVPLTW--STCA 1331

```
Db 1479 IPGLPAYILFMCVRHADYLNDDQVRSLTSTINGI-KKVLKRRGDDFETVSWLSNTCR 1537
QY 1332 FTQATENLLGDEGKPLFGALQNRHNGKALMQFAVA---QRTICPOVLIQKHLVRLLS 1388
Db 1538 F-LHCUKQSGEGGFWKHNPNRNEH-----CLTNFDLAERYVLSDLAIQIYQQLVRVLE 1592
QY 1389 VVLPNKSEDTPCLLSIDLFLVLVGAVALPPLSYWDVPLD--QPSSVS-----SSY 1438
Db 1593 NIL-----QPMIVSGMLEHETIQGSGV-----KPTGURKRTSSIADEGYTLDSII 1639
QY 1439 NHLYLPHLTMAH 1451
Db 1640 ROLNSPHSYMCQH 1652

RESULT 11
REST_HUMAN
ID REST_HUMAN STANDARD; PRT; 1427 AA.
AC P30622;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Restin (Cytoplasmic linker protein-170 alpha-2) (CLIP-170) (Reed-
DE Sternberg intermediate filament associated protein).
GN RSN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Peripheral blood monocytes;
RX MEDLINE=92289675; PubMed=1600942;
RA Bilbe G., Delabie J., Brueggem J., Richener H., Asselbergs F.A.M.,
RA Cerletti N., Sorg C., Odink K., Tarcsay L., Wiesendanger W.,
RA de Wolf-peeters C., Shipman R.;
RT "Restin: a novel intermediate filament-associated protein highly
RT expressed in the Reed-Sternberg cells of Hodgkin's disease.";
RL EMBO J. 11:2103-2113(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92405160; PubMed=1356075;
RA Pierre P., Scheel J., Rickard J.E., Kreis T.E.;
RT "CLIP-170 links endocytic vesicles to microtubules.";
RL Cell 70:1887-900(1992).
CC -!- FUNCTION: SEEMS TO BE A INTERMEDIATE FILAMENT ASSOCIATED PROTEIN
CC THAT LINKS ENDOCYTIC VESICLES TO MICROTUBULES.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC, ASSOCIATED WITH THE
CC CYTOSKELETON.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=Long;
CC IsoId=P30622-1; Sequence=Displayed;
CC Name=Short;
CC IsoId=P30622-2; Sequence=VSP_000765;
CC -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN THE REED-STERNBERG CELLS
CC OF HODGKIN'S DISEASE.
CC -!- SIMILARITY: Contains 2 CAP-Gly domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; X64838; CAA46050.1; -
CC EMBL; M97501; AAA35693.1; -
CC PIR; S22695; S22695.
CC Genew; HGNC:10461; RSN.
CC MIM; 179838; -
CC GO; GO:0005768; C:endosome; TAS.
```

```
DR GO; GO:0005882; C:intermediate filament; TAS.
DR GO; GO:0015630; C:microtubule cytoskeleton; TAS.
DR GO; GO:0008017; F:microtubule binding activity; TAS.
DR GO; GO:0006899; P:non-selective vesicle transport; TAS.
DR InterPro; IPR000938; CAP-Gly.
DR InterPro; IPR001878; Znf.CCHC.
DR Pfam; PF01302; CAP_GLY; 2.
DR SMART; SM00343; Znf.C2HC; 1.
DR PROSITE; PS00845; CAP_GLY_1; 2.
DR PROSITE; PS00845; CAP_GLY_2; 2.
DR Cytoskeleton; Microtubules; Coiled coil; Repeat; Alternative splicing.
KW DOMAIN 78 120
FT DOMAIN 143 204
FT DOMAIN 232 274
FT DOMAIN 304 331
FT DOMAIN 350 1342
FT DOMAIN 1408 1421
FT VARSPLIC 457 491
FT Missing (in isoform Short).
FT /FTId=VSP_000765.
FT CONFLICT 1069 1069 D -> E (in REF. 2).
SQ SEQUENCE 1427 AA; 0A4F16DD94254E8 CRC64;
Query Match 1.9%; Score 175.5; DB 1; Length 1427;
Best Local Similarity 19.4%; Pred No. 0.011;
Matches 200; Conservative 160; Mismatches 344; Indels 329; Gaps 45;
QY 172 EPGRAGTIKENSRCPLNEEVIVQARKIFPSVIKYVVENTIWEERKELPPELQ-----IREK 227
Db 488 EDTRVATVSEKSR-----IMELEKDLALRVQVEAELERRR 521
QY 228 NERYCYLVFNDEHSHYDHYIYSLQALDC-----ELAAEQLHTTAIDKEGRRVAKGA 280
Db 522 LESNKPAGDVDMSLSLQEISSLQEKLEVTRTDHQREITSLKEHFGARETHQKEIK-L 580
QY 281 YAAQCQ---EAKEDIKS-----HSENVQHPHLHVEVLSHSEIMAHQKFAALRLG-SWNKIMS 331
Db 581 YTATEKLSKENESLSKLEHANKENSVDIALWKSLETAIAHQOAMELKVSFSGKLG 640
QY 332 YSSDFRQIFCQ-----ACLRPEPSENPCLSRLMLDAKLYKGARKILHELI 379
Db 641 ETAEFAELKTQIEKMLDYOHEIENLQNDQSER----- 674
QY 380 FSSFFMEMEYKFLAMEEVKYKQLOKEYISDDHRSISITALSVMFTVPTLARHLIEE 439
Db 675 -AAHAKEMEAALRAKMLKVIK-EKENSLEAIRSKLDAED-----OHLVEM 717
QY 440 QNVISVITET-----LLEVLPEYLDNRNKF--NFOGY---SODKLGVRVAVICDLKYILI 489
Db 718 EDTLKLQEAIEIKVKEVLEQLAKCNEQTKVIDNFTSQLKATEKL-----LDL----- 765
QY 490 SKPTITWRLRMQFLEGRSFLKILTCMQGMEIRROVQGHIEVDPDWEA--AIAIQMOL 547
Db 766 -----DALRKAASSEG-KSEKKL--RQOLEAAEKQI-KHLEIKNAESSKASITREL 814
QY 548 KNILLMPQEWACDEELLVAYKECHKAVMRCSTSFSSKTVVQSCGSHSTKSYRSE 607
Db 815 QG-----RELKLTNLQE-----NLSEVSQVKETLEKQLILKEKFAEASE 854
QY 608 DLVSIHPLSRTLAGLHVRLSRLGAVSRLEHFVSFEDQVEVLYEPLRCLVLVAQVVAE 667
Db 855 EAVSVQRSMQETVKNLKHQKEQFNMLSS-----DLEKL----- 887
QY 668 MWRRNGLSLSQVYYQD-----VKCREEMVDKDI--IMLQIGASLMDPNKFLLLVLQR 719
Db 888 ---RENLADEAKPREKDEREEQLIKAKEKL-ENDIAEIMKMSGDNSQLTK----- 935
QY 720 YELAAEFNKTTSTKDQDLIKOYNTLIEMLQVLIYIGVGRVPGVGNVTKEVVTMEITH 779
Db 936 -----MNDELRLKERDV-----EELQLKLTKANEN--ASFLOKSIEDMTVK----- 974
QY 780 LLCIEPMPHSAIAKNLPENENNETGLENVINKVATFKKPGVSGHGVYELKDSLKDFNMY 839
Db 975 ---AEQSQEQAARKHEEKKELERKLSLEKKMET-----SHNQOQLKAR----- 1017
```





[illegible]

[1] SEQUENCE FROM N.A.  
RP MEDLINE=93024922; PubMed=1406971;  
RX Yen T.J., Li G., Schaar B.T., Szilak I., Cleveland D.W.;  
RT "CENP-E is a putative kinetochore motor that accumulates just before  
RT mitosis";  
RL Nature 359:536-539(1992).  
[2]  
RN CHARACTERIZATION.  
RP MEDLINE=95196755; PubMed=7889940;  
RX Thrower D.A., Jordan M.A., Schaar B.T., Yen T.J., Wilson L.;  
RT "Mitotic Hela cells contain a CENP-E-associated minus end-directed  
RT microtubule motor";  
RL EMBO J. 14:918-926(1995).  
[3]  
RN CHARACTERIZATION.  
RP MEDLINE=98437347; PubMed=9763420;  
RX Chan G.K.T., Schaar B.T., Yen T.J.;  
RT "Characterization of the kinetochore binding domain of CENP-E reveals  
RT interactions with the kinetochore proteins CENP-F and hBUBR1.";  
RL J. Cell Biol. 143:49-63(1998).  
CC -!- FUNCTION: MINUS-END DIRECTED MICROTUBULE MOTOR. PROBABLE  
CC KINETOCORE MOTOR. ACCUMULATES JUST BEFORE MITOSIS AT THE G2 PHASE  
CC OF THE CELL CYCLE. PROBABLY IMPORTANT FOR CHROMOSOME MOVEMENT  
CC AND/OR SPINDLE ELONGATION.  
CC -!- SUBUNIT: INTERACTS WITH CENP-F AND BUBR1 KINASE.  
CC -!- SUBCELLULAR LOCATION: ASSOCIATES WITH KINETOCHORES DURING  
CC CONGRESSION, RELOCATES TO THE SPINDLE MIDZONE AT ANAPHASE, AND IS  
CC QUANTITATIVELY DISCARDED AT THE END OF THE CELL DIVISION.  
CC -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; Z15005; CAA78727.1; -;  
DR PIR; S28261; S28261.  
DR HSPP; P17119; 3KAR.  
DR Genew; HGNC:1856; CENPE.  
DR GK; Q02224; -;  
DR MM; I17143; -;  
DR DR; GO:0005699; C:Kinetochore; TAS.  
DR DR; GO:0005634; C:nucleus; TAS.  
DR DR; GO:0008350; F:Kinetochore motor activity; TAS.  
DR DR; GO:0008067; P:DNA replication and chromosome cycle; TAS.  
DR DR; GO:0007079; P:mitotic chromosome movement; TAS.  
DR DR; GO:0007080; P:mitotic metaphase plate congression; TAS.  
DR DR; InterPro; IPR001752; kinesin\_motor.  
DR DR; Pfam; PF00225; kinesin; 1.  
DR DR; PRINTS; PR00380; KINESINHEAVY.  
DR DR; SMART; SM00129; KISC; 1.  
DR DR; PROSITE; PS00411; KINESIN\_MOTOR\_DOMAIN1; 1.  
DR DR; PROSITE; PS50067; KINESIN\_MOTOR\_DOMAIN2; 1.  
KW Motor protein; Cell division; ATP-binding; Coiled coil; Mitosis;  
KW Cell cycle; Centromere.  
FT DOMAIN 1 335 KINESIN-MOTOR.  
FT DOMAIN 336 2471 COILED COIL (POTENTIAL).  
FT DOMAIN 2472 2663 GLOBULAR (POTENTIAL).  
FT NP\_BIND 86 93 ATP (BY SIMILARITY).  
SQ SEQUENCE 2663 AA; 312087 MW; CEFC13880C8C8CB8 CRC64;  
  
Query Match 1.9%; Score 175.5; DB 1; Length 2663;  
Best Local Similarity 19.4%; Pred. No. 0.028;  
Matches 228; Conservative 189; Mismatches 406; Indels 355; Gaps 55;  
  
QY 179 IKNSRCPLNEEVIVQARKIFPSVIKYVEMTWEEKEKLPPEL-----QIREKNERYYC 233  
Db 733 VEENE--ALPEEVI-----LLSELKSLPSVEVRLKREIODKSEEL-- 770

```

Query Match      1.9%; Score 175.5; DB 1; Length 2663;
Best Local Similarity 19.4%; Pred. No. 0.028;
Matches 228; Conservative 189; Mismatches 406; Indels 355; Gaps 55;

Qy 179 IKNSRCPLNEEVIVQARKIFPSVIVVEMTWEEEEKLPPEL-----QIREKNERYVC 233
      :|||
      :|||

Db 733 VEENE--ALREEV-----LLSEKSLPSVERLRKEIQDSEEL-- 770

```



use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL: AJ237839; CAB41415.1; -  
 EMBL: AF000692; BAA89431.1; -  
 EMBL: AF000689; BAA89431.1; JOINED.  
 EMBL: AF000690; BAA89431.1; JOINED.  
 EMBL: AF000691; BAA89431.1; JOINED.  
 EMBL: AF001725; BAA95548.1; -  
 EMBL: AB023150; BAA76777.2; -  
 Genew; HGNC:1291; C21orf5.  
 MIM; 604803; -  
 GO; GO:0007275; P:development; NAS.  
 Pfam; PF04118; Dopey\_N; 1.  
 CONFLICT 498 498 Q -> H (IN REF. 1).  
 CONFLICT 1118 1118 H -> C (IN REF. 2 AND 3).  
 CONFLICT 1149 1149 H -> P (IN REF. 2 AND 3).  
 CONFLICT 1318 1318 L -> P (IN REF. 1).  
 CONFLICT 1886 1886 V -> A (IN REF. 1).  
 CONFLICT 1974 1974 E -> G (IN REF. 1).  
 CONFLICT 2139 2139 G -> E (IN REF. 4).  
 CONFLICT 2154 2154 K -> R (IN REF. 1).  
 SEQUENCE 2298 AA; 258221 MW; 2C449AFE550739F0 CRC64;

Query Match 1.9%; Score 170; DB 1; Length 2298;

Best Local Similarity 18.3%; Pred. No. 0.05;  
 Matches 288; Conservative 201; Mismatches 551; Indels 536; Gaps 70;

QY 122 IDPTCVLCMDFQDSVHKHRYKMTSTGGFCDCGDTAEW----KTG--PFCVNEPGR 175  
 DB LEPVILLQ-----PKTQTSIHCLKQENRAD--DLHRWNRKTSFREACAVPEQE 1047  
 QY 176 AGTIKNSRCPLEEVIVQARKIPFVSVIKYVEMTIWEEKEKLPELQIREKNERYCVL 235  
 DB EELPLSQFTVDREALWAEVK-----EPEKPYLRGELSEELPYVEL 1095  
 QY 236 FNDEHSHDVHYISLQALDCELAELAHHTAIDKEGRRAVAGAAOEAEDKSHS 295  
 DB PORTAG-----APDSSEHTSADTSSGH-----TDS 1122  
 QY 296 ENVSHPLHVEVLHSEIMAHQFALRLGSMNKNINSYSDFRQIFCQACLRPEPSNCP 355  
 DB ENTS-----SFSPSHD-----LQELSNENCC 1145  
 QY 356 LISRLMLDAKLYKGARKILHELIFSSFFEMEYKKLFAMEFVKY--YKOLQKEYISDDH 413  
 DB -----APTHMGGRAPKRSALLAFAQSFSEFKAGAKLSLVYVDSKDTQASESFSSDE 1196  
 QY 414 DRSISITALSVQMTPTPTLARHLIEBQNVISVITETLLEVLPEYLDNRNKNFNGYSQDK 473  
 DB EADLELQAL-----TTSRLKQOREQEAVALFKHILLYLQPYDSR----- 1238  
 QY 474 LGRVAVICDLKYLISKPTIWTERRMQFLGRFSFLKILTCMQMEEIRROVGOHIEV 533  
 DB --RVLYAFSVLEAVLKNP-----KEFIEAVSRSTSDTSTAHLNLSINLLARH--- 1285  
 QY 534 DPDWEAATA-----IQOLKNILLMFQEWACDEELLV-----AYKEC--- 572  
 DB ---QALIGQSYGKLGQTQVNV-----CPHSLLELLTYLCLSLFSLRSPYCYLKV 1333  
 QY 573 -HKAYMRCSTSPISSSKTVVQ-----SCGHSLETKSYRVSEDLVS-----IHLPLSR 618  
 DB SHRDILNRDQVQKSVLELIRIMMQLVSVAKSSECKNVEFIHSLQRCVKQOEVLVLS- 1392  
 QY 619 TLAGHLVRLSRIGAVSRHL-----EFVSFEDE-----QVEVLVEYPLRC-LVLVAQV--- 664  
 DB --ASMYTSQKRYGLATAHGRALPEDSLPEESLINLGQDIWSEHPLQIELLKLLQVLIV 1450  
 QY 665 -----VAEMWR--NGLSLISQVFFYQD-----VKCREEM 692  
 DB LEHLGRAHEEAENOPDLRSREWRALNFOQAISALQYVQPHPLTSQGLLVSAYVRGLQPA 1510

QY 693 YDKDI-----IMLOIGASLMDPNKPELLLLVLQRYELAEAFN 727  
 DB YGYCMHPAWVSLVTHSLDPYFKSGLWTVTFVQVQCKNLDD-----LVKOYE-SESVK 1562  
 QY 728 KTIST--KDQDLIKOYN--TLIEEMLQVLIYIVGERYVPGVGNVTKEEVTMRRIIHLCLIE 784  
 DB LSVSTTSKRENISPDYPLTLEGLTTISHFCLLEQ-----ANQNKTTAAGDPANL--- 1613  
 QY 785 PMPHSATAKNLPENENNETGLENVINKVATPKKPGVSGHGVYELKDESLKDNMMFYHVS 844  
 DB RNARNAILEELPRTVNTMALLWNVLKREETOKRP-----VDLLGATKGSSSVYF-KTT 1665  
 QY 845 KTOHSHK-----AEHMOKKRKOENKDBALPPPPPEF---CP 878  
 DB KTIQKILDFLNPHTAHLGVOLTAAVAAWKSKAQRHSMKILIPTASASQLTLDVLC- 1724  
 QY 879 AFSKVINLLCNDIMMYILRTVFER-----AIDTDSNLWTEGLMAFHIL----- 923  
 DB ALSTLOTDTLLHLVKVEVKKRPPQVQGGDEKSLVDIPVLQFCYAFQLRPLVPALQE 1780  
 QY 924 ----ALGLEEKQOLQKAPEEVTFDYHKASRLGSSAMNIQMLEKLKIPQLEQKDM 979  
 DB NFSLLGVKESVQLNLAPP-----GYFLLSMLNDEVTTRTPNLENKKDQ 1825  
 QY 980 ITWILQMFDTVKRLREKSLIVATSGSESIKNDIETHDKKAEKRKAARLHRQKTM 1039  
 DB KDLQETQKILEAVNIAGSS-----L 1847  
 QY 1040 AQMSALOKNFTEHKLMYDNTSEMPGKED-----SIMEEESTPAVSDYSRIALGPKRG 1092  
 DB EQTSWLSRN-LEVKAQPAQASLEESDAEEDLDYDAAAASAMVSSASVSVQALSL----- 1901  
 QY 1093 PSVTEKEVLTCIL-----COEQEVKTEN-----NAMVLSACVQKSTA 1130  
 DB LAEVLASLLDMVYRSDEKAVPLISRLLYVYVFLRNHSAYNAPSFRAGALLSS 1957  
 QY 1131 LTQHRGPIELSGBALDPLFMDPLAYCTYTGSCGHVHMAVCQKYFEAVQLSSQRIHV 1190  
 DB LSGYAYTKRAWRKVLE-LFLDP--AFFQMDTSCVH-----WKSIIIDL-LTHEKTMFK 2007  
 QY 1191 DLFPLESG-----EVLCPCLKSLCNTVIP-----IIPLOKIN-----SENAD 1229  
 DB DLMNQSSSKLFSFQKAMLLRQAPAFSGELDQYHLYLPLQIERTDMLRGQTSI 2067  
 QY 1230 ALAQLLTLARIQVTLARISGYNIRHAKGNPIPT-----FNGQMGDSTLEFHSILSFG 1284  
 DB VAAQMFLEFR--VLLLRISP--QHLTSLMFWIMVSELIQFTQLEEDLKDEDESURSTN 2121  
 QY 1285 VESSIKYS-----NSIKE-----MVLFTATTYIRIGLVKPPDPRVPMLTWSTCAF 1332  
 DB KVNRTKVSVPDANGPSVGEIPOSELILYLSACKFLDTALSFPDP---KMLPFQIYVWAF 2177  
 QY 1333 TIQAIENLGDGKPLFGALONRONGKALKMQFAVAORITCPQVLOKHILVRLISV-- 1390  
 DB IPEY-----DTEGPAFLSDVEENHQEC-----PHTVRILELKL 2212  
 QY 1391 -LPNIKSEDPTCLLSIDLFLHVLGAVLAFPSLYWDPDVLOPSSVSSVSNHLYLPHLITM 1449  
 DB KFEISSSEITMKS-----EFPL-----LRQHSVSSIRQLMPFFWTLMG 2252  
 QY 1450 AHMLQILLVDT-GLP 1464  
 DB AFKTQRLPADSPGTP 2268

RESULT 15

AKA9 HUMAN

ID AKA9\_HUMAN STANDARD; PRT: 3911 AA.  
 AC Q99956; O14869; O43355; O94895; Q9UQH3; Q9UQ04; Q9Y6B8; Q9Y6Y2;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE A-kinase anchor protein 9 (Protein kinase A anchoring protein 9)  
DE (PRK9) (A-kinase anchor protein 450 kDa) (AKAP 450) (A-kinase anchor  
DE protein 350 kDa) (AKAP 350) (ngAKAP 350) (AKAP 120 like protein)  
DE (hyperion protein) (Yotiao protein) (Centrosome- and Golgi-localized  
DE PN-associated protein) (CG-NAP).  
GN AKAP9 OR AKAP450 OR AKAP350 OR KIAA0803.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RN SEQUENCE FROM N.A. (ISOFORM 4).  
RP TISSUE=Brain;  
RX MEDLINE=98151389; PubMed=9482789;  
RA Lin J.W., Wyszynski M., Madhavan R., Sealock R., Kim J.U., Sheng M.;  
RT "Yotiao, a novel protein of neuromuscular junction and brain that  
RT interacts with specific splice variants of NMDA receptor subunit  
RT Nr1.1";  
RL J. Neurosci. 18:2017-2027(1998).  
RN [2]  
RN SEQUENCE FROM N.A. (ISOFORM 2), AND VARIANT GLN-1347 INS.  
RX MEDLINE=99219864; PubMed=10202149;  
RA Witczak O., Skaalhegg B.S., Keryer G., Bornens M., Tasken K.,  
RA Jahnsen T., Oerstavik S.;  
RT "Cloning and characterization of a cDNA encoding an A-kinase anchoring  
RT protein located in the centrosome, AKAP450.";  
RL EMBO J. 18:1858-1868(1999).  
RN [3]  
RN SEQUENCE FROM N.A. (ISOFORM 3).  
RP TISSUE=Brain;  
RX MEDLINE=99287934; PubMed=10358086;  
RA Takahashi M., Shibata H., Shimakawa M., Miyamoto M., Mukai H., Ono Y.;  
RT "Characterization of a novel giant scaffolding protein, CG-NAP, that  
RT anchors multiple signaling enzymes to centrosome and the Golgi  
RT apparatus.";  
RL J. Biol. Chem. 274:17267-17274(1999).  
RN [4]  
RN SEQUENCE FROM N.A. (ISOFORM 1).  
RA Kemmer W.A., Deiss S., Schwarz U.;  
RT "Cloning of Hyperion.";  
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RN SEQUENCE OF 323-3911 FROM N.A. (ISOFORM 2).  
RP TISSUE=Gastric parietal cell;  
RX MEDLINE=99115654; PubMed=9915845;  
RA Schmidt P.H., Dransfield D.T., Claudio J.O., Hawley R.G.,  
RA Trotter K.W., Milgram S.L., Goldenring J.R.;  
RT "AKAP350, a multiply spliced protein kinase A-anchoring protein  
RT associated with centrosomes.";  
RL J. Biol. Chem. 274:3055-3066(1999).  
RN [6]  
RN SEQUENCE OF 1802-3876 FROM N.A. (ISOFORM 5).  
RP TISSUE=Lymphoblast;  
RA Hinds K., Sutterer C., Becker M., Hawkins M.;  
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.  
RN [7]  
RN SEQUENCE OF 2157-3911 FROM N.A. (ISOFORM 6).  
RP TISSUE=Lung;  
RA Milgram S.L., Goldenring J.R., Schmidt P.H.;  
RT "AKAP350: A multiply spliced family of proteins with centrosomal  
RT association.";  
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.  
RN [8]  
RN SEQUENCE OF 2212-3911 FROM N.A. (ISOFORM 2/3).  
RP TISSUE=Brain;  
RX MEDLINE=99087487; PubMed=9872452;  
RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Miyajima N.,  
RA Tanaka A., Kotani H., Nomura N., Ohara O.;  
RT "Prediction of the coding sequences of unidentified human genes. XI.  
RT The complete sequences of 100 new cDNA clones from brain which code  
RT for large proteins in vitro.";  
RL DNA Res. 5:277-286(1998).  
RN [9]

RP SEQUENCE OF 17-1800 FROM N.A.  
RA Wu X., Graves T., Bradshaw H.;  
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: BINDS TO TYPE II REGULATORY SUBUNITS OF PROTEIN KINASE  
CC A. SCAFFOLDING PROTEIN THAT ASSEMBLES SEVERAL PROTEIN KINASES AND  
CC PHOSPHATASES ON CENTROSOME AND GOLGI APPARATUS WHERE PHYSIOLOGICAL  
CC EVENTS CAN BE REGULATED BY PHOSPHORYLATION STATE OF PROTEIN  
CC SUBSTRATES. ISOFORM 4/YOTIAO IS ASSOCIATED WITH THE N-METHYL-D-  
CC ASPARTATE RECEPTOR AND IS SPECIFICALLY FOUND IN THE NEUROMUSCULAR  
CC JUNCTION (NMJ) AS WELL AS IN NEURONAL SYNAPSES EXPLAINING THAT ITS  
CC ROLE MAY BE TO ORGANIZE POSTSYNAPTIC SPECIALIZATIONS.  
CC -!- SUBUNIT: INTERACTS WITH THE REGULATORY REGION OF PROTEIN KINASE N  
CC (PKN). PROTEIN PHOSPHATASE 2A (PP2A). PROTEIN PHOSPHATASE 1 (PPI)  
CC AND THE IMMATURE NON-PHOSPHORYLATED FORM OF PKC EPSILON.  
CC -!- SUBCELLULAR LOCATION: CENTROSOmal IN MANY CELL TYPES AND  
CC CYTOPLASMIC IN PARIETAL CELLS.  
CC -!- ALTERNATIVE PRODUCTS:  
CC Event-Alternative splicing; Named isoforms=6;  
CC Name=1;  
CC IsoId=Q99996-1; Sequence=Displayed;  
CC Name=2;  
CC IsoId=Q99996-2; Sequence=VSP\_004102, VSP\_004107;  
CC Name=3; Synonyms=CG-NAP;  
CC IsoId=Q99996-3; Sequence=VSP\_004102, VSP\_004105, VSP\_004107;  
CC Name=4; Synonyms=Yotiao;  
CC IsoId=Q99996-4; Sequence=VSP\_004103, VSP\_004104;  
CC Name=5;  
CC IsoId=Q99996-5; Sequence=VSP\_004108;  
CC Name=6; Synonyms=AKAP350;  
CC IsoId=Q99996-6; Sequence=VSP\_004106, VSP\_004107, VSP\_004109;  
CC -!- TISSUE SPECIFICITY: WIDELY EXPRESSED. ISOFORM 4/YOTIAO IS HIGHLY  
CC EXPRESSED IN SKELETAL MUSCLE AND IN PANCREAS.  
CC -!- DOMAIN: RII BINDING SITE, PREDICTED TO FORM AN AMPHIPATHIC HELIX,  
CC COULD PARTICIPATE IN PROTEIN-PROTEIN INTERACTIONS WITH A  
CC COMPLEMENTARY SURFACE ON THE R-SUBUNIT DIMER.  
CC -!- CAUTION: REF.6 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO TWO  
CC FRAMESHIFTS IN POSITIONS 3782 AND 3811.  
CC -!- CAUTION: REF.9 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO FOUR  
CC FRAMESHIFTS IN POSITIONS 29, 1653, 1699 AND 1735.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AJ131693; CAB40713.1; -;  
DR EMBL; AB019691; BAA78718.1; -;  
DR EMBL; AJ010770; CAA09361.1; -;  
DR EMBL; AF026245; AAB86384.1; -;  
DR EMBL; AF083037; AAD22767.1; -;  
DR EMBL; AC004013; AAB96867.1; ALT\_FRAME.  
DR EMBL; AF091711; AAD39719.1; -;  
DR EMBL; AB018346; BAA34523.1; -;  
DR EMBL; AC000066; AAC60380.1; ALT\_FRAME.  
DR Genew; HGNC:379; AKAP9.  
DR MIM; 604001; -;  
DR GO; GO:0005813; C:centrosome; TAS.  
DR GO; GO:0005856; C:cytoskeleton; TAS.  
DR GO; GO:0004973; F:N-methyl-D-aspartate receptor-associated pr...; TAS.  
DR GO; GO:0005515; F:protein binding activity; TAS.  
DR GO; GO:0007165; P:signal transduction; TAS.  
DR GO; GO:0006832; P:small molecule transport; TAS.  
DR GO; GO:0007268; P:synaptic transmission; TAS.  
DR Coiled coil; Alternative splicing; Polymorphism.  
KW DOMAIN 2554 2567 PKA-RII SUBUNIT BINDING DOMAIN.  
FT DOMAIN 164 914 COILED COIL (POTENTIAL).  
FT DOMAIN 944 1022 COILED COIL (POTENTIAL).  
FT DOMAIN 1100 1185 COILED COIL (POTENTIAL).  
FT DOMAIN 1253 1280 COILED COIL (POTENTIAL).  
FT DOMAIN 1336 1392 COILED COIL (POTENTIAL).

FT	DOMAIN	1434	1459	COILED COIL (POTENTIAL).
FT	DOMAIN	1585	1659	COILED COIL (POTENTIAL).
FT	DOMAIN	1857	2455	COILED COIL (POTENTIAL).
FT	DOMAIN	2544	2561	COILED COIL (POTENTIAL).
FT	DOMAIN	2603	2776	COILED COIL (POTENTIAL).
FT	DOMAIN	3065	3092	COILED COIL (POTENTIAL).
FT	DOMAIN	3124	3470	COILED COIL (POTENTIAL).
FT	DOMAIN	3587	3689	COILED COIL (POTENTIAL).
FT	DOMAIN	3726	3730	POLY-LEU.
FT	DOMAIN	203	292	GLN-RICH.
FT	DOMAIN	321	1010	GLU-RICH.
FT	DOMAIN	1846	2772	Missing (in isoform 2 and isoform 3).
FT	VARSPLIC	17	28	/FTid-VSP_004102.
FT	VARSPLIC	1637	1642	QLOEEI -> LATRD (in isoform 4).
FT	VARSPLIC	1643	3911	/FTid-VSP_004103.
FT	VARSPLIC	2175	2182	Missing (in isoform 4).
FT	VARSPLIC	2175	2182	/FTid-VSP_004104.
FT	VARSPLIC	2175	2183	Missing (in isoform 3).
FT	VARSPLIC	2895	2907	/FTid-VSP_004105.
FT	VARSPLIC	2895	2907	SADTFQKVE -> Q (in isoform 6).
FT	VARSPLIC	2895	2907	/FTid-VSP_004106.
FT	VARSPLIC	2895	2948	VGFYNNCFSTLC -> GSSIPBLAHSDAYOTREICSS
FT	VARSPLIC	3901	3911	(in isoform 2, isoform 3 and isoform 6).
FT	VARSPLIC	3901	3911	/FTid-VSP_004107.
FT	VARSPLIC	3901	3911	Missing (in isoform 5).
FT	VARSPLIC	3901	3911	/FTid-VSP_004108.
FT	VARSPLIC	3901	3911	STQPHAGMRR -> ALSLTTSWOHSARPTAPLFFELSH
FT	VARSPLIC	3901	3911	SLG (in isoform 6).
FT	VARIANT	1347	1347	/FTid-VSP_004109.
FT	VARIANT	1347	1347	K -> KQ.
FT	CONFLICT	76	76	/FTid-VAR_010926.
FT	CONFLICT	475	475	E -> Q (IN REF. 3).
FT	CONFLICT	554	554	M -> I (IN REF. 3).
FT	CONFLICT	638	638	E -> G (IN REF. 3).
FT	CONFLICT	663	663	R -> S (IN REF. 3).
FT	CONFLICT	913	913	N -> S (IN REF. 3).
FT	CONFLICT	956	956	H -> N (IN REF. 3).
FT	CONFLICT	980	982	K -> N (IN REF. 3).
FT	CONFLICT	997	997	K -> N (IN REF. 3).
FT	CONFLICT	1001	1001	QKH -> PRP (IN REF. 1 AND 2).
FT	CONFLICT	1020	1020	Q -> P (IN REF. 1 AND 2).
FT	CONFLICT	1028	1028	Q -> P (IN REF. 1 AND 2).
FT	CONFLICT	1626	1626	V -> D (IN REF. 3).
FT	CONFLICT	1703	1703	V -> E (IN REF. 3).
FT	CONFLICT	1707	1707	R -> P (IN REF. 1 AND 2).
FT	CONFLICT	1802	1803	N -> T (IN REF. 3).
FT	CONFLICT	1843	1843	V -> G (IN REF. 3).
FT	CONFLICT	1843	1843	MISSING (IN REF. 5).
FT	CONFLICT	1843	1843	A -> P (IN REF. 3).
Query Match				1.8%; Score 168.5; DB 1; Length 3911;
Best Local Similarity				18.7%; Pred. No. 0.14;
Matches 201; Conservative 165; Mismatches 364; Indels 347; Gaps 50;				
QY	208	EMTWEKEKPELQIRKERNRYCVLFNDEHSHVDHYISLQRLDCELAQAHTTA	267	
Db	728	EMTL--QINELQKEIEILRQEEK-----EKGTEQEVEQLQ--LKTELLEKOMK---	772	
QY	268	IDKEGRRVAKAGAYACQAEKEDIKSHSENVQHPHLVHLVHSEIMAHOKFALRLGSMWN	327	
Db	773	-EKENDLQKFAQLAENSILKDEKTLLEDMLK-----IHTPVSOEERLI-----FLD	819	
QY	328	KIMSYSSDP---RQIFCQACLRPEPDSNPCLISRLMLMDAKLYKGARKILHELIFSSFF	384	
Db	820	SIKSKSDSWKEI--ELIENEDLKQCCI-----QLNSEEIKQRNTFFSAEKN	868	
QY	385	MEMEYKFLFAMFVYKOLQKEY-----ISDHDHSISITALSVQMTFTPTLARHLIEE	439	
Db	869	FEVN-----YQELQEEYACLLKVKDDLEDKSN-----KQELLEY	901	
QY	440	QNVISVITETLLELPEYLDNRN-----KFNQGYSDKLGRRYAVICDLKYLISKPTIW	495	
Db	902	RSKLKALNEEL-----HLQRINPTVTKMKSSVFDEDK--TFVAETLEMEGVVEKDTTEL	953	

QY	496	TERLRMQFLGFRSFLKILTCMQGMBEIRRVQGO-HIEVDDPDWEAAIAIATOMQLKNILLMF	554	
Db	954	MEKLEVTKREK-----LELSQRLSEQLQKHGEISFLNEEVKSLKQKEQVSLRC	1006	
QY	555	QEW-----CACDEE-----LLLVAYKECHKAVMRCSSTFSSSTKTVVQSCG	595	
Db	1007	RELEIIINHNRANENVQSCDTQVSSLLDGVVVTWSRGAESGVKNVKSFEESKIMWED--	1064	
QY	596	HSLETKSYRVSED-----LVSIHLPLSRTLGLH-VRLSRLGAVSRHLHFVSFEDQVEV	649	
Db	1065	-KVSFENNTVGEESKQEQILDLHLSVTKESSLRATQPSENDKLQKLNVLKSEQNDLRL	1123	
QY	650	LVEPLRCLVLVAQVVAEMWRNRGLSLISQVYVODVKCREMYDKDIIMLQIGASLMDP	709	
Db	1124	QMEAQRICLSLVYS-----THVDQVREYME-----NEKDKALCSLKEELIFA	1165	
QY	710	NKELLVLQR-YELAAEFNKTIISTKDQD-----LTKOYNTLIEE-----MLOVLIYIVGER	759	
Db	1166	QEEKIKELQKIHQLELQTMKTQETGDEGKPLHLIGLKQKAVSEBSCSYFLQTLCSVLGEY	1225	
QY	760	YVPGVGNVTKBEVTMRIIHLICIEBPMPSALAKNLPENNETGLENVINKVATEKKPG	819	
Db	1226	YTPAL-----KCEV-----NAEDKEN-----	1241	
QY	820	VSGHGVVELKDESUKDFNMVYHYHYSKTOHSAEHMOKRRKROENKDEALPPPPPPFCPA	879	
Db	1242	-SGDYISENDEPELDQYRYEQDFQENMHT-----LLNKVTEEYNK-----	1281	
QY	880	FSKVINLLNCIDIMVILRTVFERAIDTDSNLW---TEGMLQMAFHILALGLLEEKQOLQK	936	
Db	1282	-----LLVLQTRL-----SKIWQOOTDGM-----KLEFGE-----N	1308	
QY	937	APREEVTF-----DFYHKASRLGSSAMNIOMLLKLGIPQLEGOKDMITWI	983	
Db	1309	LPKEETEFLSHSQMTNLEDIDVNHK-----SKLSSLODLEKTKLEEQVQLESLSL	1362	
QY	984	LQMFDTVKRLREKS--CL-----IVATTSGSESIKNDITHDKKAERKKAFAARLHRQ	1036	
Db	1363	QOOLKETEQNYEAETHCLQKRLQAVSESTVPPSLPVDVSVITESDA-----Q	1409	
QY	1037	KIMAQMSALQKNFTETHKLMYDNTSEMPECKEDSIMEEESTPAVSDYSRIALGPKRGPSVT	1096	
Db	1410	RTMYPGSCVKKNI-----DGTIEFSG-EFGVKEETNI-----VKLL	1444	
QY	1097	EKEVLTCILCOEVEVKIENAMVLSACVQKSTALTQHRGKPIELSGEALDPLFMDPDLA	1156	
Db	1445	EKQY-----QEQLLEEVAKVIVSVMSIAFAQOTELSR-----ISG-----	1478	
QY	1157	YGYTGTSGGHVYMHAVCWOK--YFEAVQLSSQORIHVDLF---DLESGEYLCPLCKSL	1208	
Db	1479	-GKENTASSKQAHAVCQEQEHYFNEMKL-SQDQIGFQTFETVDVVKFKEEFKPLSKEL	1533	

Search completed: September 25, 2003, 14:55:22  
Job time : 50 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 25, 2003, 14:51:51 ; Search time 148 Seconds  
(without alignments)  
3023.400 Million cell updates/sec

Title: US-09-724-126A-19  
Perfect score: 9141  
Sequence: 1 AMEGNMADEAGGTERMEIS.....ETARSQETNQMIFGFNWQLL 1734

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_23.\*  
1: sp\_archaea.\*  
2: sp\_bacteria.\*  
3: sp\_fungi.\*  
4: sp\_human.\*  
5: sp\_invertebrate.\*  
6: sp\_mammal.\*  
7: sp\_mhc.\*  
8: sp\_organelle.\*  
9: sp\_phage.\*  
10: sp\_plant.\*  
11: sp\_rodent.\*  
12: sp\_virus.\*  
13: sp\_vertebrate.\*  
14: sp\_unclassified.\*  
15: sp\_rvirus.\*  
16: sp\_bacteriap.\*  
17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	9087	99.4	1749	4 Q8IYW7	Q8IYW7 homo sapien
2	8791.5	96.2	1709	4 Q8IYW6	Q8IYW6 homo sapien
3	8480	92.8	1757	11 Q70481	Q70481 mus musculus
4	4262	46.6	849	11 Q8BN40	Q8BN40 mus musculus
5	4229	46.3	1755	4 Q8IYW8	Q8IYW8 homo sapien
6	4209	46.0	811	4 Q96JY4	Q96JY4 homo sapien
7	4184	45.8	1755	11 Q8CGW0	Q8CGW0 mus musculus
8	3062	33.5	1275	4 Q15057	Q15057 homo sapien
9	3053	33.4	1273	4 Q9NU68	Q9NU68 homo sapien
10	2593.5	28.4	1109	11 Q8K2I6	Q8K2I6 mus musculus
11	2232.5	24.4	861	11 Q8BUL9	Q8BUL9 mus musculus
12	2210.5	24.2	1824	5 Q8SX71	Q8SX71 drosophila
13	2208.5	24.2	1824	5 Q9VX91	Q9VX91 drosophila
14	1886.5	20.6	1927	5 P91133	P91133 caenorhabdi
15	1716	18.8	333	4 Q60708	Q60708 homo sapien
16	1692	18.5	329	4 Q75492	Q75492 homo sapien

17	1688	18.5	777	11 Q8ROV7	Q8ROV7 mus musculus
18	1067	11.7	229	11 Q8C5K3	Q8C5K3 mus musculus
19	911.5	10.0	410	11 Q8RI30	Q8RI30 mus musculus
20	606.5	6.6	1225	10 Q9LZ95	Q9LZ95 arabidopsis
21	598	6.5	212	4 Q9H578	Q9H578 homo sapien
22	535.5	5.9	1872	3 Q07963	Q07963 saccharomyc
23	513	5.6	2153	5 Q19330	Q19330 caenorhabdi
24	470.5	5.1	1264	10 Q94JL2	Q94JL2 oryza sativ
25	362	4.0	1470	5 Q9W3M3	Q9W3M3 drosophila
26	359	3.9	2230	5 Q8NMN0	Q8NMN0 dictyosteli
27	304	3.3	731	5 Q9W3M5	Q9W3M5 drosophila
28	303	3.3	1432	5 Q8IRN8	Q8IRN8 drosophila
29	292	3.2	795	10 Q9LZ94	Q9LZ94 arabidopsis
30	288.5	3.2	5439	5 Q8I586	Q8I586 plasmodium
31	284.5	3.1	794	10 Q96248	Q96248 arabidopsis
32	257	2.8	437	4 Q8IWE7	Q8IWE7 homo sapien
33	252	2.8	343	4 Q8ND96	Q8ND96 homo sapien
34	208	2.3	378	11 Q8CAW0	Q8CAW0 mus musculus
35	205.5	2.2	3187	11 Q63714	Q63714 rattus norv
36	196	2.1	280	10 Q94JL1	Q94JL1 oryza sativ
37	189.5	2.1	8749	4 Q8NF91	Q8NF91 homo sapien
38	186	2.0	1410	4 Q14221	Q14221 homo sapien
39	185	2.0	1449	6 Q9BG87	Q9BG87 bos taurus
40	184	2.0	1411	4 Q15075	Q15075 homo sapien
41	184	2.0	1979	4 Q95949	Q95949 homo sapien
42	182.5	2.0	2442	4 Q9H450	Q9H450 homo sapien
43	182	2.0	2442	4 Q14812	Q14812 homo sapien
44	181.5	2.0	2442	4 Q60588	Q60588 homo sapien
45	181.5	2.0	2579	11 Q9ESD3	Q9ESD3 mus musculus

#### ALIGNMENTS

#### RESULT 1

Q8IYW7 PRELIMINARY; PRT; 1749 AA.  
ID Q8IYW7  
AC Q8IYW7;  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Ubiquitin ligase E3 alpha-I.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Han H.Q., Kwak K.;  
RT "Full-length human ubiquitin ligase E3 alpha-I (E3 alpha).";  
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AY061886; AAL32103.1;  
KW Ligase.  
SQ SEQUENCE 1749 AA; 200210 MW; 3AE0E1A749884971 CRC64;

Query Match	99.4%;	Score 9087;	DB 4;	Length 1749;
Best Local Similarity	98.8%;	Pred. No. 0;		
Matches 1728;	Conservative	0;	Mismatches	1; Indels 20; Gaps 1;
Qy	6	MADEEAGGTERMEISAEELPOTPPQRLASWMDQVDFYTAFLHHLAQLVPEIYFAEMDPDL 65		
Db	1	MADEEAGGTERMEISAEELPOTPPQRLASWMDQVDFYTAFLHHLAQLVPEIYFAEMDPDL 60		
Qy	66	KQESVQMSFTPLEWYLFGEPPDPCLEKHKHSAFQLCGRVFKSGEYTCRDCADPT 125		
Db	61	KQESVQMSFTPLEWYLFGEPPDPCLEKHKHSAFQLCGRVFKSGEYTCRDCADPT 120		
Qy	126	CVLCMDCQFQSVKHNHRYKMHSTGGGFCDCGDTFAWKTPFCVNHPEGRAGTIKENSRC 185		
Db	121	CVLCMDCQFQSVKHNHRYKMHSTGGGFCDCGDTFAWKTPFCVNHPEGRAGTIKENSRC 180		
Qy	186	PLNEEVTVQARKIPSPSVIKYVWETIWEELPELPQIREKKNERYVCVLFNDEHSHSDH 245		

Db 181 PLNEEVIVQARKIPPSVIKYVWMTIWEKEKELPELOIREKNERYCYVLENDHSHSYDH 240  
Qy 246 VIYSLQALDCELAEOHLHTAIDKEGRRAVAGAYAACOPAKEDIKSHSNVSOHPLHV 305  
Db 241 VIYSLQALDCELAEOHLHTAIDKEGRRAVAGAYAACOPAKEDIKSHSNVSOHPLHV 300  
Qy 306 EVLHSEITMAHOKFALRLGSMWNTMSYSSDFRQIFCOACLEEDPDSENPCLISRLMLWA 365  
Db 301 EVLHSEITMAHOKFALRLGSMWNTMSYSSDFRQIFCOACLEEDPDSENPCLISRLMLWA 360  
Qy 366 KLYGARKILHELLFSFFMEMEYKILFAMEFVYKQLOKEYISDDHDSISITALSQV 425  
Db 361 KLYGARKILHELLFSFFMEMEYKILFAMEFVYKQLOKEYISDDHDSISITALSQV 420  
Qy 426 MFTVPTLARHLIEBQNVISVITETLLEVLPEYLDNRNKNFNQGYSDQKLGRIYAVICDLK 485  
Db 421 MFTVPTLARHLIEBQNVISVITETLLEVLPEYLDNRNKNFNQGYSDQKLGRIYAVICDLK 480  
Qy 486 YILISKPTIWTERRUMQFLEGRFSLKILTCMQGMEIRROVQGHIEVDPDWEAAIAIQM 545  
Db 481 YILISKPTIWTERRUMQFLEGRFSLKILTCMQGMEIRROVQGHIEVDPDWEAAIAIQM 540  
Qy 546 QLKNIILMFQWCACDELLIYAYKECHKAVMRCSTSFISSTKTVQSCGHSLETKSYRV 605  
Db 541 QLKNIILMFQWCACDELLIYAYKECHKAVMRCSTSFISSTKTVQSCGHSLETKSYRV 600  
Qy 606 SEDLVSITLPLSRTLAGLHVRLSLGAVSRILHEFVSFEDFQVEVLVEYPLRCLVLVAQV 665  
Db 601 SEDLVSITLPLSRTLAGLHVRLSLGAVSRILHEFVSFEDFQVEVLVEYPLRCLVLVAQV 660  
Qy 666 AEMWRRNGLSLSQVFFYQDYKCREMYDKDIIMLOIGASLMDPNKFLLLVLQRYELAE 725  
Db 661 AEMWRRNGLSLSQVFFYQDYKCREMYDKDIIMLOIGASLMDPNKFLLLVLQRYELAE 720  
Qy 726 FNKTIISTKQDOLIKOYNTLIBEMLOVLIYIVGERYVPGVGNVTKKEVTMREIHLHCIEP 785  
Db 721 FNKTIISTKQDOLIKOYNTLIBEMLOVLIYIVGERYVPGVGNVTKKEVTMREIHLHCIEP 780  
Qy 786 MPHSAIAKNLPENNETGLENVINKVATFKPKPGVSGHGVYELKDESLKDFNMVPHYSK 845  
Db 781 MPHSAIAKNLPENNETGLENVINKVATFKPKPGVSGHGVYELKDESLKDFNMVPHYSK 840  
Qy 846 TQSKAEHMQKRRKQENKDALPPPPPPPCPAFSKVINLLNCDIMMYILRTVPERAID 905  
Db 841 TQSKAEHMQKRRKQENKDALPPPPPPPCPAFSKVINLLNCDIMMYILRTVPERAID 900  
Qy 906 TDSNLWTGMLQMAFHIALGELLEKEQOLOKAPPEEVTDFYHKASRLGSSAMNTOMLLE 965  
Db 901 TDSNLWTGMLQMAFHIALGELLEKEQOLOKAPPEEVTDFYHKASRLGSSAMNTOMLLE 960  
Qy 966 KLKGIPQLEGOKDMITWLOMFDVTKRLRKSCLIVATTSGSESIKNDEITHDKKABRK 1025  
Db 961 KLKGIPQLEGOKDMITWLOMFDVTKRLRKSCLIVATTSGSESIKNDEITHDKKABRK 1020  
Qy 1026 RKAEARLHROKIMQMSALOKNFETHKLMYDNTSEMPGKEDSTIMEESTPAVSDYSRI 1085  
Db 1021 RKAEARLHROKIMQMSALOKNFETHKLMYDNTSEMPGKEDSTIMEESTPAVSDYSRI 1080  
Qy 1086 ALGPKRGPSTVEKEVLTICLQEOEVEKIENAMVLSACVOKSTALTOHRGKPIELSGEA 1145  
Db 1081 ALGPKRGPSTVEKEVLTICLQEOEVEKIENAMVLSACVOKSTALTOHRGKPIELSGEA 1140  
Qy 1146 LDPLFMDPDLAGYTGTCGSHVMHACVQKFEAVQLSSQORIHVDLFDLESSEYLCPLC 1205  
Db 1141 LDPLFMDPDLAGYTGTCGSHVMHACVQKFEAVQLSSQORIHVDLFDLESSEYLCPLC 1200  
Qy 1206 KSLCNTVPIIPILOPQKINSNADALQLLTLARMIQTVLARISGYNIRHAKGNPIPIF 1265  
Db 1201 KSLCNTVPIIPILOPQKINSNADALQLLTLARMIQTVLARISGYNIRHAKGNPIPIF 1260  
Qy 1266 FNOGMDSTLBFHSILTSFGVSSIKYSIKEMVTLFATTIYRIGLKVPPDPRVPM 1325  
Db 1261 FNOGMDSTLBFHSILTSFGVSSIKYSIKEMVTLFATTIYRIGLKVPPDPRVPM 1320

Qy 1326 TWSTCAFTIQTATENLLGDEGKPLFGALQNRQHNGLKALMQFAVAQORITCPQVLIQKHLVR 1385  
Db 1321 TWSTCAFTIQTATENLLGDEGKPLFGALQNRQHNGLKALMQFAVAQORITCPQVLIQKHLVR 1380  
Qy 1386 LLSVVLPNIKSBDTPCLLSIDLFLHVLGAVLAPPSLYWDDPVDLPQSSVSSYNHLYLFH 1445  
Db 1381 LLSVVLPNIKSBDTPCLLSIDLFLHVLGAVLAPPSLYWDDPVDLPQSSVSSYNHLYLFH 1440  
Qy 1446 LITMAHMLQILLTVDITGLTGLAQVQEDSEEAHSASSFFAEISQYTSIGCDIPGWYLMWS 1505  
Db 1441 LITMAHMLQILLTVDITGLTGLAQVQEDSEEAHSASSFFAEISQYTSIGCDIPGWYLMWS 1500  
Qy 1506 LKNGITPYLRCAALFFHYLLGVTPPEELHTNSAEGYSALCSYLSLPTNLFLLFQEWYWD 1565  
Db 1501 LKNGITPYLRCAALFFHYLLGVTPPEELHTNSAEGYSALCSYLSLPTNLFLLFQEWYWD 1560  
Qy 1566 VRPILORRCADPALLNCLKOKNTVVRYPKRNSLIJELPDDYSCLLNQASHFRCPRSADDE 1625  
Db 1561 VRPILORRCADPALLNCLKOKNTVVRYPKRNSLIJELPDDYSCLLNQASHFRCPRSADDE 1620  
Qy 1626 RKHPVLCIFCGAILCSQICCOEIVNGEEVGCIFHALHC----- 1665  
Db 1621 RKHPVLCIFCGAILCSQICCOEIVNGEEVGCIFHALHC----- 1660  
Qy 1666 KARGCAYPAPYDDEYGETDGLKRGPNLHLSRERYRKLHLVWQOHCIEEARSQETNQM 1725  
Db 1681 KARGCAYPAPYDDEYGETDGLKRGPNLHLSRERYRKLHLVWQOHCIEEARSQETNQM 1740  
Qy 1726 LFGFNWQLL 1734  
Db 1741 LFGFNWQLL 1749

RESULT 2  
Q8IWY6 PRELIMINARY; PRT; 1709 AA.  
ID Q8IWY6 PRELIMINARY;  
AC Q8IWY6;  
DT 01-MAR-2003 (Tremblrel. 23, Created)  
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)  
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
DE UB1 E3a ligase (Fragment).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=22340442; PubMed=12434312;  
RA Dgany O., Avidan N., Delaunay J., Krasnov T., Shalmon L., Shalev H.,  
RA Eidelitz-Markus T., Kapelushnik J., Cattani D., Pariente A.,  
RA Tulliez M., Cretien A., Schischmanoff P.O., Tolascio A., Fibach E.,  
RA Koren A., Rossler J., Le Meirer M., Yaniv I., Zaizov R., Ben-Asher E.,  
RA Olender T., Lancet D., Beckmann J.S., Tamary H.;  
RT "Congenital dyserythropoietic anemia type I is caused by mutations in  
RT codanin-1";  
RL Am. J. Hum. Genet. 71:1467-1474 (2002).  
DR EMBL; AF525401; AA014997.1; --  
KW Ligase.  
FT NON\_TER 1 1709  
FT NON\_TER 1709 1709  
SQ SEQUENCE 1709 AA; 195279 MW; B2B17231A1020FID CRC64;

Query Match 96.2%; Score 8791.5; DB 4; Length 1709;  
Best Local Similarity 98.0%; Pred. No. 0;  
Matches 1676; Conservative 1; Mismatches 12; Indels 21; Gaps 2;

Qy 23 LPQTPORLASWMDQVDFYTAFLHHLAQLVPEYFAEMDPLEKEQESVQMSIFTPLEWY 82  
Db 1 LPQTPORLASWMDQVDFYTAFLHHLAQLVPEYFAEMDPLEKEQESVQMSIFTPLEWY 60  
Qy 83 LFGEDPDICLEKLKHSAGFQLCGRVPKSGETTSYCRDCAIDPTCVLCMDCFDQSVKHNHR 142



Db 61 LFGEEDPICLEKLKHSFAFQLGGRVFKSGGETTYSRCDCAIDPTCVLCMDCFQDSVHKNR 120  
QY 143 YKMHSTGCGFCGDTAEAWKTCPCVNHPEGRAGTIKENSRCPLNEEVIVQARKFPSPV 202  
Db 121 YKMHSTGCGFCGDTAEAWKTCPCVNHPEGRAGTIKENSRCPLNEEVIVQARKFPSPV 180  
QY 203 IKYVYEMTWEBEKELPPELQIREKNERYCYVLFNDEHHSYDHVYISLQALDCELAEAQ 262  
Db 181 IKYVYEMTWEBEKELPPELQIREKNERYCYVLFNDEHHSYDHVYISLQALDCELAEAQ 240  
QY 263 LHTTAIDREGRRAVAGAYAAQOEAKEDIKSHSENVSQHPLHVEVHSEIMAHQKFAIRL 322  
Db 241 LHTTAIDREGRRAVAGAYAAQOEAKEDIKSHSENVSQHPLHVEVHSEIMAHQKFAIRL 300  
QY 323 GSWMNKINSYSDFRQIFCOACLRPEPDSFNCLISRLMDAKLYKGARKIILHELIFSS 382  
Db 301 GSWMNKINSYSDFRQIFCOACLRPEPDSFNCLISRLMDAKLYKGARKIILHELIFSS 360  
QY 383 FFMENYKFLFAMFVKYKQLOKEYISDDHRSISITALSVMQFTVPTLARHLIEQNV 442  
Db 361 FFMENYKFLFAMFVKYKQLOKEYISDDHRSISITALSVMQFTVPTLARHLIEQNV 420  
QY 443 ISVITETLLEVPYLDLDRNKNFQGYSDQKLGVRVAVICDLKYIILISKPTIWTERRMQ 502  
Db 421 ISVITETLLEVPYLDLDRNKNFQGYSDQKLGVRVAVICDLKYIILISKPTIWTERRMQ 480  
QY 503 FLEGFRSFLKIULTQMGMEIIRQVQGHTEVPDWEAAIAIOMQLKNILLMFQEWACADE 562  
Db 481 FLEGFRSFLKIULTQMGMEIIRQVQGHTEVPDWEAAIAIOMQLKNILLMFQEWACADE 540  
QY 563 ELLLVAYKECHKAVMRCSTFSISSKTVVQSGHSLSTKSYRVEDLVSIHLPLSRTLAG 622  
Db 541 ELLLVAYKECHKAVMRCSTFSISSKTVVQSGHSLSTKSYRVEDLVSIHLPLSRTLAG 600  
QY 623 LHVRLSRLGAVSRLEHFEVSFDFQVEVLVEYPLRCLVLVAQVVAEMWRNGLSLISQVYF 682  
Db 601 LHVRLSRLGAVSRLEHFEVSFDFQVEVLVEYPLRCLVLVAQVVAEMWRNGLSLISQVYF 660  
QY 683 YQDVKCREMYDKDIIMIQIGASLMDPNKFLLLVLRQYELAEAFNKTISTKQDOLIKQYN 742  
Db 661 YQDVKCREMYDKDIIMIQIGASLMDPNKFLLLVLRQYELAEAFNKTISTKQDOLIKQYN 720  
QY 743 TLIEEMQLVLIYIGERYVPGVNTKERTVWREIHLICIEPMPHSAIAKNLPENENNE 802  
Db 721 TLIEEMQLVLIYIGERYVPGVNTKERTVWREIHLICIEPMPHSAIAKNLPENENNE 780  
QY 803 TGLENVINKVATFKPGVSGHGVYELKDLSKDFNMVYHYSKTOHSAEHMOKRRKOE 862  
Db 781 TGLENVINKVATFKPGVSGHGVYELKDLSKDFNMVYHYSKTOHSAEHMOKRRKOE 840  
QY 863 NKDEALPPPPPEFCPAFSKVINLNCIDIMYILRTVFERAIDTDSNLTWTEGMLQWAFHI 922  
Db 841 NKDEALPPPPPEFCPAFSKVINLNCIDIMYILRTVFERAIDTDSNLTWTEGMLQWAFHI 900  
QY 923 LALGLLEKQQLQKAPEBEVTFDFYHKASRLGSSAMNIQMLLEKLGIPQLEGQKDMITW 982  
Db 901 LALGLLEKQQLQKAPEBEVTFDFYHKASRLGSSAMNIQMLLEKLGIPQLEGQKDMITW 960  
QY 983 ILOMFDTVKRLREKSLIVATTSGESIKNDIETHDKAEKRAEAAARLHROKIMAO 1042  
Db 961 ILOMFDTVKRLREKSLIVATTSGESIKNDIETHDKAEKRAEAAARLHROKIMAO 1020  
QY 1043 SALQKNFTETHKLYMDNTSEMPGKEDSIWEESTPAVSDYSRIALGPKRGPSVTEKEVLT 1102  
Db 1021 SALQKNFTETHKLYMDNTSEMPGKEDSIWEESTPAVSDYSRIALGPKRGPSVTEKEVLT 1080  
QY 1103 CILCQEEQEVKIENNAMVLSACVQKSTALTQHRGKPIELSGEALDPLFMDPDLAYGTYTG 1162  
Db 1081 CILCQEEQEVKIENNAMVLSACVQKSTALTQHRGKPIELSGEALDPLFMDPDLAYGTYTG 1140  
QY 1163 SCGHVMAHVCWKQFEAVQLSSQQRHIVDLFDESGEVILCPICKSLCNTVPIIPLQPOK 1222  
Db 1141 SCGHVMAHVCWKQFEAVQLSSQQRHIVDLFDESGEVILCPICKSLCNTVPIIPLQPOK 1200

QY 1223 INSENADALAOQLLTLARWIOQVLARISGYNIRHAKGENPIPIFFNQMGDSTLEPHSILS 1282  
Db 1201 INSENADALAOQLLTLARWIOQVLARISGYNIRHAKGENPIPIFFNQMGDSTLEPHSILS 1260  
QY 1283 FGVESSIKYSIKSEMVLIFATTIYRIGLKVPPDERDPRVPMLTWSTCAFTIOAILENLG 1342  
Db 1261 FGVESSIKYSIKSEMVLIFATTIYRIGLKVPPDERDPRVPMLTWSTCAFTIOAILENLG 1320  
QY 1343 DEGKPLFGALQNRHNGLKALMQFAVAQRITCPQVLIQKHLVRLLSVVLPNIKSEDTPL 1402  
Db 1321 DEGKPLFGALQNRHNGLKALMQFAVAQRITCPQVLIQKHLVRLLSVVLPNIKSEDTPL 1380  
QY 1403 LSIDLFLVLCVLAFLPSLYWDDPVDLPQSSVSSSYNHLFLHLITMAHMLQILLTVDTG 1462  
Db 1381 LSIDLFLVLCVLAFLPSLYWDDPVDLPQSSVSSSYNHLFLHLITMAHMLQILLTVDTG 1440  
QY 1463 LPLAQOVEDSEEAHSASSFFAEISQYTSIGSDIPGWYLMVSLKNGITPYURCAALRPH 1522  
Db 1441 LPLAQOVEDSEEAHSASSFFAEISQYTSIGSDIPGWYLMVSLKNGITPYURCAALRPH 1500  
QY 1523 YLLGVTPEELHTNSARGEYSALCSYLSLPTNLFLFOEYWDVTRPLLRRCADPALLNC 1582  
Db 1501 YLLGVTPEELHTNSARGEYSALCSYLSLPTNLFLFOEYWDVTRPLLRRCADPALLNC 1560  
QY 1583 LKOKNTVVRYPKRNSLIELPDDYSCLLNQASHFCRPSADDERKHPVLCFCGAILCSQ 1642  
Db 1561 LKOKNTVVRYPKRNSLIELPDDYSCLLNQASHFCRPSADDERKHPVLCFCGAILCSQ 1620  
QY 1643 NICCOEIVNGEEVGCACIFHALHC-----KARGCAYPAPYLDEYGE 1682  
Db 1621 NICCOEIVNGEEVGCACIFHALHCAGVCIFLIRECRVVLVEGKARGCAYPAPYLDEYGE 1680  
QY 1683 TDPLGKRGNPLHLSRERYRKLHLVWQOHC 1712  
Db 1681 TDPLGKRGNPLHLSRERYRKLHLVWQOHC 1709  
RESULT 3  
070481 PRELIMINARY; PRT: 1757 AA.  
AC 070481  
DT 01-AUG-1998 (TrEMBLrel. 07, Created)  
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Ubiquitin-protein ligase E3 COMPONENT N-recognin (Ubiquitin-protein  
ligase E3-alpha).  
GN UBRI.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Kwon Y.T., Reiss Y., Fried V.A., Hershko A., Yoon J.K., Gonda D.K.,  
RA Sangon P., Copeland N.G., Jenkins N.A., Varshavsky A.;  
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE OF 1-446 FROM N.A.  
RC STRAIN=129/SVJ;  
RA Kwon Y.T., Reiss Y., Fried V.A., Hershko A., Yoon J.K., Gonda D.K.,  
RA Sangon P., Copeland N.G., Jenkins N.A., Varshavsky A.;  
RT "The mouse and human genes encoding the recognition component of the  
RT N-end rule pathway.";  
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF061555; AAC40165.1; -  
DR EMBL; AF067379; AAC23678.1; -  
DR EMBL; AF067371; AAC23678.1; JOINED.  
DR EMBL; AF067372; AAC23678.1; JOINED.  
DR EMBL; AF067373; AAC23678.1; JOINED.  
DR EMBL; AF067374; AAC23678.1; JOINED.  
DR EMBL; AF067375; AAC23678.1; JOINED.  
DR EMBL; AF067376; AAC23678.1; JOINED.







```
Db 481 REEMYDKDIIMLQIGASLMDPNKFLLLVLQRYELAEAFNKTISTKDQDLIKQYNTLIEEM 540
QY 749 LQVLIYVGERYVPGVGNVTKEVTRWEIILHLCIEPMPSATAKNLPENNETGLENV 808
Db 541 LQVLIYVGERYVPGVGNVTKEVTRWEIILHLCIEPMPSATAKNLPENNETGLENV 600
QY 809 INKVAFTFKPGVSGHGYELKDSKDFNMNFYHYSKTQSHKAHMQKRRKQENKDEAL 868
Db 601 INKVAFTFKPGVSGHGYELKDSKDFNMNFYHYSKTQSHKAHMQKRRKQENKDEAL 660
QY 869 PPPPPPEFCAPSKVNLNLCNDIMMILRTVFERAIDTDSNLTWEGMLQMAFHIALGLL 928
Db 661 PPPPPPEFCAPSKVNLNLCNDIMMILRTVFERAIDTDSNLTWEGMLQMAFHIALGLL 720
QY 929 EKKQOLQKAPEEVTDFYHKSRLGSSAMNIOMLLEKLGIPOLGOKDMITWLOMFD 988
Db 721 EKKQOLQKAPEEVTDFYHKSRLGSSAMNIOMLLEKLGIPOLGOKDMITWLOMFD 780
QY 989 TVKRLREKSLIVATTSGSESKNDIEITHDK 1019
Db 781 TVKRLREKSLIVATTSGSESKNDIEITHDK 811

RESULT 7
Q8CGW0 PRELIMINARY; PRT; 1755 AA.
AC Q8CGW0;
DT 01-MAR-2003 (Tremblrel. 23, Created)
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Ubiquitin ligase E3 alpha-II (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C3H/HeN;
RA Han H.Q., Kwak K.;
RT "Novel ubiquitin ligase E3 alpha-II.";
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY061885; AAL32102.1; -.
KW Ligase.
FT NON_TER
SQ SEQUENCE 1755 1755
Query Match 45.8%; Score 4184; DB 11; Length 1755;
Best Local Similarity 45.9%; Pred. No. 9.5e-283;
Matches 824; Conservative 335; Mismatches 531; Indels 104; Gaps 29;

QY 6 MADE---EAGGTER--METSALPQTPQRLASWQDQDFYTAFLHHLAQLVPEIYFAEM 60
Db 1 MASEMEPEVQADIRSLLECSAE-----EIAGRWLQATDLNREYVQHIAHCVPKIYCRGP 54
QY 61 DPDLKQERSVOMSTFTPLEWILFGEDDPDICELEKXHSG-AFOLGGRVFKSGTITVYSCRD 119
Db 55 NFPPOKEDTLAGHLLGPNWEVICAEDPALGFPKLEQANKPSHLCGRVFKVGEPTYSCRD 114
QY 120 CAIDPTCVLMDCFODSVHKNHRYKMHSTGTGGGFCDCGDTGTEAKWTGPFVCVNHPEGRAGTI 179
Db 115 CAVDPTCVLCMCFGLSHRDHRYRMTTSGGGFCDCGDTGTEAKWKGYPYCKHKLSSSEVV 174
QY 180 -KENSRCPLNEEVIQVQARKIPFVSVIKYVVMETIWEKEKLPPLQIREKNERYCYVLFND 238
Db 175 EEDPLVHLSEDVIARTYNIFAIMFYAVDILTWEKESELPEDEVAEKSDTYCYMLFND 234
QY 239 EHHSDHYVILQSORDCLABELAQLHTTAIDKEGRRAVKAGAYAAQCOAEKEDIKSHENV 298
Db 235 EVHTTEQVIYTLQKAVNCTQKEAIGFATTVDRDRRPVRYGDFQYCDQAKTVIVRNTSRQ 294
QY 299 SOHPLHVEVLHSEIMAHQKFAIRLGSWMNMKIMSYSDRFQICQACLREPDSENCLIS 358
Db 295 TK-PLKVQVMHSSVAHQFGLKALSWLGSVIGYSDGLRILLCQVGLQCGPDGENSLVD 353
```

```
QY 359 RLMLWDAKLYKGARKILHELIFSSFFMEMEYKKLFAMEFVKYKQLOKQYISDDHDSIS 418
Db 354 RLMLNDSKLWAGARSYYHQLFMSSLLMDLYKKLFAKFNKYLQLODFMEDDHERAYS 413
QY 419 ITALSVQMTPTTLARHLIEONVISTETLEVLPEYLDNRNKNFNFQGS---QDKLG 475
Db 414 VTALSVQFTTAPTALRMLLTENLMTVIKAFMDHL-KHRDAQGRQFERYTALQAFKFR 472
QY 476 RYAVIVIDCKYLISKPTIWTIRLQMQFLEGRSFLKILTKMQGMBEIRRQVQGHIEVDP 535
Db 473 RVQSLILDLYKVLISKPTESWDELQKFLQGFDALELLKCMQGMDDPITRQVQGHITEMP 532
QY 536 DWEAIAIQMOLKNILLMFOEWCADDELLLVAYKECHKAVMRCSTSFSSSKTYVQS-C 594
Db 533 EWEAAFTLQMLKLTHTVISMVQDWCALDEKVLIEAYKKCLAVLTQCHGFTDGGQPTLSIC 592
QY 595 GHSLETYSYRVSEDLVSIHLPLSRTLGLHVLRLSRLGAVSRSLHEFVSFFDFQVEVLVEYP 654
Db 593 GHSVETIRYCVSQEKVSIHLPLSRLLAGLHVLLSKSEVAYKFPPELLPLSELPPMLIEHP 652
QY 655 LRLVLVAQVVAEMWRRNGLSLISQVFFYQDVVKREMYDKDIIMLQIGASLMDPNKFL 714
Db 653 LRLVLCAQVHAGWRRNGFSLVNGIYYVHNVKCRREMFKDKDIVMLQTVGSMMDPNHFLM 712
QY 715 LVLQRYELAEAFN-----KTISTK--DQDLIKQYNTLIEEMLOVLIYIVGERYVPGVNV 767
Db 713 IMLSRFELYQLFSTPDYKGRFSSEVTHKDVVQNNNTLIEEMLYLIIMLVGERFNPQGVQ 772
QY 768 TKEBVTWTEIILHLCIEPMPSATAKNLPENNETGLENVINKVATFKKPGVSGHGYE 827
Db 773 AATDEIKRELIHQLSIKPMASSELVKSLPDENKETGMEVSIVAHFKKPGVTRGMYE 832
QY 828 LKDESLKDFNMFFYHYSKTQSHKAHMQKRRKQENKDEALPPPPPPPCPAFASKVINLL 887
Db 833 LKPECAKEFNLYFYHFSRAEQSKAEAEQKLRKENKEDTALPPALPPCPLFASLVNLL 892
QY 888 NCDIMMYLILRTVFERAIDTDSNLTWEGMLQMAFHIALGLLEKQOLQKAPBEV--TFDF 946
Db 893 QCDVNLXIMGITLQWAVEHHGSAWSESLQRYLHLIGMALQEKHLENAVEGHVQTFTF 952
QY 947 YHKASRLGSSAMN---IQMLLEKLGIPOLGOKDMITWLOMFTVLRLEKSKSLIVAT 1003
Db 953 TQKISKPGDAPHNPSILAMLETQNASPLSAHKMDIRWLLKMFNAIKKIRE--CSSSP 1010
QY 1004 TSGSESKNDIEITHKEKAERKAEARLRQKIMQASALQKNFIETHKLMYDNTSEM 1063
Db 1011 VAEAGTIMESSRDKAKERKAEIARLRREKIMQASEMQRHFDENKELFOOTLEL 1070
QY 1064 PGKEDSIMEESTPAVSYSRIALGPKRGPSTVEKEVLTICLQBEQEVYKIENNAMVLSA 1123
Db 1071 DTSASATL--DSSPPVSDAALTALGPAQTQVPEPRQFVTCILCQEQEVTVGSRAMVLA 1128
QY 1124 CVQKSTALTQHRGPIELSGEALDPLFMDPOLAYGTYTSCGHVHVHVCWKYFPAVQLS 1183
Db 1129 FVQRSTVLSKDRTKTI-ADPEKYDPLFMHPDLSCGTHTSGCGHVHVAHCWQRYFSVQAK 1187
QY 1184 SQQ-----RHVDLFDLESGEVLCPLCKSLCNTVPIPILOPKINSENADALAULTLA 1238
Db 1188 EQRQQRRLHTS-YDVENGFEFLPCLECLNTVPIPL-LPPRSILSRRLN-FSQOPDLA 1244
QY 1239 RWIOTVLARISGYNI---RHAKGEN-----PIPIFFNMGMDSTLEFHSILSF 1283
Db 1245 QWTRAVTQQIKVVQMLRRKKNAAADTSSSESTEAMNIIPPEGRFP-----DFY---- 1292
QY 1284 GVSESIKYSNIKEMVILFATTIYRIGLKVPPDERDPVPMLTWTSTCAFTIQAIBNLGSD 1343
Db 1293 ---PRNPYSIKEMLTFTFGTAAYKVLKHPNEGDPVPLCMWCTCAYTIQISIERILSD 1349
QY 1344 EGRPLFCALQNRQHNGKALMQFAVAQRITCPQVLIQKHLVRLLSVLPNKSDETPCLL 1403
Db 1350 EEKPVFGPLPRLCDCLRSLTRFAAAHWTVALLPVVOGHFKFLFASLPVSDSYEDLPCL 1409
```

```

QY 1404 SIDFLVHLVAVLAPSLYWDPPVDLPQSSVSSSYNHLILFLHLITMAHMLQIILL--TVD 1460
D 1410 DIDFHLVGLVLAFFALQCD-----FSGSSLATGDLHIFHLVTHMAHIVQIILTSCTEE 1464
QY 1461 TGLPLAQVQEDSEEAHSASSFAEISQYTSIGSDICDIPGWVLSLKNIGTPYLRCAALF 1520
D 1465 NGMD--QENPTGEBELAILSLHKTHTQYTSALKEAPSGWHLWSVRAAIMPFLKCSALF 1522
QY 1521 FHYLGVTPPELHTNSAEGEYSALCSYLSLPTNLFLLFOEYWTYRPLRQRCADPALL 1580
D 1523 FHYLGVPAAPPDQV--SGTSHFELCNLYSLPTNLHLFQENSDIMNSLIESWCQNSVK 1581
QY 1581 NCLKQKTVVYPRKRSNLIPLDDYSCLLNQASHFRCPRSADDERKHPVLCFCGAILC 1640
D 1582 RYLNGERGAIISYPRGANKLIDLDPEDYSLLNQASNFSCPKSGGDKSRAPTICLCVGSLLC 1641
QY 1641 SONTCCQEIYNGEVGACIFALHC-----KARGCAYPAPYLDEY 1680
D 1642 QSVCQAELGEDEVGACTAHTYSCGSGAGIFLRVRECOVLFLAGTKGCSYSPYLDY 1701
QY 1681 GETDPGLKRNPLHLRSRYRKYRLHLVWQHCIIIEIARSOETNOMLFGFNQQL 1734
D 1702 GETDGLRRGNPLHLQCEFRKIKLQWQHSITEIGHAQEANTLVGIDWQHL 1755

RESULT 8
O15057 PRELIMINARY; PRT; 1275 AA.
AC O15057;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein KIAA0349 (Fragment).
GN KIAA0349.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=97349984; PubMed=9205841;
RA Nagase T., Ishikawa K., Nakajima D., Ohira M., Seki N., Miyajima N.,
RA Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. VII.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.";
RL DNA Res. 4:141-150(1997).
DR EMBL; AB002347; BAA20806.1; -.
DR InterPro; IPR000408; Reg_chromatins.
DR SMART; SM001841; Znf_ring.
DR SMART; SM00184; RING; 1.
DR PROSITE; PS00626; RCCL_2; 1.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 1275 AA; 101FF1F66E056066 CRC64;

Query Match 33.5%; Score 3062; DB 4; Length 1275;
Best Local Similarity 46.4%; Pred. No. 1.3e-204;
Matches 600; Conservative 242; Mismatches 391; Indels 60; Gaps 20;

QY 484 LKYLISKPTITWRLRMQLEGRFSFLKILTCMQMGEIRROVGOHIEVDPDWEAAIAI 543
D 1 KYVLIISKPTWSEDLKQKFLLEGDFDAFLKLCMQGMDPTIRQVGOHLEMEPEWAAFTL 60
QY 544 QMOLKNILMFQEWACDEELLYAYKECHKAVMRCSTSFSSSKTVVQS--CGHSLETKS 602
D 61 QMKLTHVISMWQDCASDEKVLIEAYKKLAVLMQCHGYTDGEQPTILSICGHSVEIR 120
QY 603 YRVSEDLVSLHPLSRTLAGLHVLRLSGLAVSRHLHEFVSFDFQVEVLVEPLRLCLVLA 662
D 121 YCVSQEKVSTHPLVSRLLAGLHLLSKSEVAIKPPELLPLSELSPMLLIEHPLRLCLVLA 180

```

```

QY 663 QVVAEMWRNGLSLISQVYFYODVKCREMYDKDIIMLOIGASLMDPNKFLLLVLRVEL 722
D 181 QVHAGMWRNRGSLVQNYIYYHNKCRREMFDDVVMQGTGTSMMDPNKFLIMLSRREL 240
QY 723 AEAFN-----DKTIK--DQDLIKQNTLIEMLQVLIIVIGERYVPGVNTKEBVTMR 775
D 241 YQIFSTPDYGRFSSEITHKDVVQNNLTIEEMLYLIIMLVGERSPGVQVGNATDEIR 300
QY 776 EIIHLCLTEPMPSAIAKNPENNETGLEVINNKVATFKKPGVSGHGVYELKDSLDK 835
D 301 EIIHOLSIPMAHSELVKSLPDENKETGMSVIEAVAHEFKKPGTGRGMVELKPECAKE 360
QY 836 FNMYYHYHYSKTOHSAEHMQKRRKQENKDEALPPPPPPPCFAPAFSKVINLLNCIDIMYI 895
D 361 FNLHYFHSRAEQSAEEAQRKLQRNREDTALPPVLPFCPLFASLNILOSVDYMLCI 420
QY 896 LRTVPERAIDTNSLWTEGMLQMAFHILALGLEEKQOLQKAPEE--VTFDFVHKASRLG 954
D 421 MGTILQWAVEHNGYAWSESMLQRLVHLIGMALQOEKQHLNTEEHVVTFTTQKLSKPG 480
QY 955 SSAMN---IOMLEKLGIPQLEGOKDMITWILQMFDTVKRLREKS--CLIVATTSGESI 1010
D 481 EAPKNSPSILAMLETQNAPLYEVHKDMIRWILKTFNAVKKRESPTSPVAETECTIM- 539
QY 1011 KNDETHDKAEARRKAEARLHRQKIMASALQKNFIETHKLMYDNTSEMPGKEDSI 1070
D 540 --EESRDKDKAERKRAEIAIRLRREKIMASSEMQRHFDENKELFOOTLELDASTAV 597
QY 1071 MEEESTPAVDSYRIALGPKRGPSYTEKEVLTCILCOEEOEYKIENNAVLSACVOKSTA 1130
D 598 LDH--SPVASDMTLTALGTQVPEQRFVTCILCOEEOEYKVESRAVLAAFVORSTV 655
QY 1131 LTOHRGKPIELSGEALDPLFMDPLAYGTYTSCGHVMHACVQKYEAVQLSSQQ--- 1186
D 656 LSKNRSKFIQ--DPEKYDPLFMPHDLSCGTHTSSCGHIMAHACWQRFDSVQAKEQRQR 714
QY 1187 -RIHVDLFDLESGEYLCPLCKSLCNTVPIIPILOPKQINSENADALQALLTLARWITQVL 1245
D 715 LRLHTS--YDVENGFEFLCPLCECLSNVTIPL--LPPRNIFNNRLN--FSDQPNLTQWIRTIS 771
QY 1246 ARISGYNIRHAKGENPIPIFFNQMGCDSTLEPHSLISFGVESIKYSNKIKENAVILPAT 1305
D 772 QOIKALQFLRKEESTPNNAKTNSENVDELQLOPEGRFDPFRPKIPYSESIKEMLTFTFTA 831
QY 1306 IYRIGLKVPPDRDPRVPMLTWSTCAFTIQTENLGLDGEKPLFGALQNRHNGKALMQ 1365
D 832 TYKVLKLVHPNEEDPRVPMCWGSCAYTIQSIERILSDEKPLFGPLPCRLDDCLSLTR 891
QY 1366 FAVAQRITCPQVLIQKHLVRLSLVPLNPKISEDTPCLLSIDLPHVLVGAVALPSPSYWDD 1425
D 892 FAAAAHTVASVSVQGHFKFLFASLVPNDSHEELPCILDTIDMFHLLVGLVLAPALQCD 951
QY 1426 PVDLPSSVSSSYNHLILFLHLITMAHMLQIILL--TVDTGLPLAQVQEDSEEAHSASSFF 1482
D 952 -----FSGISLGTGDLHIFHLVTHMAHIIQIILTSCTEENGMD--QENPCEESAVLALY 1004
QY 1483 AEISQYTSIGSDICDIP--GWLAVSLKNGITPYLRCAALFHLVLLGYTPPELHTNSAEGE 1541
D 1005 KTLHYIT--GSAKEIPSGWHLWSRVAGIMPFLKCSALFHYLHLLVSPSPDIOV--PGTSH 1062
QY 1542 YSALCSYLSLPTNLFLLFOEYWDTVRPLLRQRCADPALLNCLKOKNTVYRPRKRSNLI 1601
D 1063 FEHLCSYLSLPLNNLCLFQENSEIMNSLIESWCNRSEVKRYLBERDAIRYPRESNKLIN 1122
QY 1602 LPDDYSCLLNQASHFRCPRSADDERKHPVLCFCGAILCSQNTCCQEIYNGEVGACIFH 1661
D 1123 LPEDYSSLLNQASNFSCPKSGGDKSRAPTICLCVGSLLCSQSYCCQTELEGEDVGACTAH 1182
QY 1662 ALHC-----KARGCAYPAPYLDEYCEYDTPDKLGRGNPLHLRSRYR 1701
D 1183 TVCSGSGVGIPLVRQCVLFLAGTKGCSYSPYLDYDVTGTQOGLRRGNPLHLRERFK 1242
QY 1702 KLHLVWQHCIIIEIARSOETNOMLFGFNQQL 1734

```







```

539  EVTHKDVQONNTLIEEMLYLIMLGERENPGVGVAATDEIKREIIHQLSJKPMHSE 598
791  IAKNLPENNETGLENVINKVATFKKPGVSGHGVYKDESLKDFNMVYHYKTOHKS 850
599  LVKSLPEDENKETSIVESAHEFKKPGLTGRGMVELPECAKEFNLYFYHFSRAQSK 658
851  AEHMOKKRRKQENKDALPPPPPCPAFSKVINLLNCDIMMYILRTVFERAIDTNSL 910
659  ABEAQKLRKREKNDTALPPALPPCPFLFASLVNLTQCDVMYIMGTILQWAVEHHGSA 718
911  WTEGMLQAFHIALGLLEKQOLAKAPEEV--TFDYHKASRLGSSAMN--IQMLEK 966
719  WESMLQRLVHLGLMAQEKHLENAVEGHVGTFTTQKISKPGDAPHNSPSILAMLET 778
967  LRGIPLEGOKDMITWILQDFVTKRLREKSLIVATTSGSEIKNDEITHDKEAKRR 1026
779  LQNAPSLEAKHOMIRLLKMFNAIKIRE--CSSSPVAEAGTMEESSRDKKAERK 836
1027 KAEARLHRQKMAQMSALQKNPTE 1051
837  KAEIARLRREKIMAOFMORHFD 861

RESULT 12
Q8SX71
ID Q8SX71 PRELIMINARY; PRT; 1824 AA.
AC Q8SX71;
DT 01-JUN-2002 (trEMBLrel. 21, Created)
DT 01-JUN-2002 (trEMBLrel. 21, Last sequence update)
DT 01-NAR-2003 (trEMBLrel. 23, Last annotation update)
DE LD31957p.
OS CG9086.
GN Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
RA Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,
RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA Celiker S.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY094815; AAM11168.1; -.
DR FlyBase; FBgn030809; CG9086.
DR InterPro; IPR003126; znf_Nrecognin.
DR InterPro; IPR001841; znf_ring.
DR Pfam; PF02207; zf-UBR1; 1.
DR SMART; SM00184; RING; 1.
DR SMART; SM00396; znf_UBR1; 1.
SQ SEQUENCE 1824 AA; 208329 MW; D68FCB81D35C7124 CRC64;

Query Match 24.2%; Score 2210.5; DB 5; Length 1824;
Best Local Similarity 31.1%; Pred. No. 5.9e-145;
Matches 596; Conservative 300; Mismatches 700; Indels 323; Gaps 56;

QY 10 EAGGTERMEISALPOTPORLASWDDQVDFTAFLLHLLAQVPEIYFAEMDPDLKQEE 69
DB 29 QAGTLDRSDIIIEFLKRESPK---YFDYQT---SATVKDNTVITLCKMFKE---SLAKEE- 78
QY 70 SYQMSITFTPLEWLFGEEDPICLEKLHSG-AFQLCGRVFKSGETYSRCDCAIDPTCVL 128
DB 79 -----IIDVVVEFLGDNPSALEKRLLEGNTATVCGKFKNGEPTYSRCGVDPTCVL 133
QY 129 CMDCFQDSVHKHRYKMHMTSTGGGFCDCGDTBAWKTGPFCVNHPEGRAGTIKENSRCPLN 188
DB 134 CVNCFKRSARHPKHYKMTSGGGGCGDCGDEAWKDKQYCELHLANKNPLESKI---LT 190
```

```

189  EEVIVQARKIFPSVIKYVEMTIWEEEXELPPELOIREKN-----ERYVCVLFNDEH 240
191  DAVLERVEICGAILAFCSYL-----ETEPNASLQCLDGNVEGGQVDGAQYCTVYNDES 246
241  HSYDHVYISLORALDCELAEAQLHTAJDKEGRRRAVKAGAYAAACOAEAKEDIKSH----- 294
247  HTFQVQIOTLTKIAKCRAKDAWEIVAAIDREGRAVVKCDTPECNKLLKYSIENQMLPTPS 306
295  --SENVSOPHLHVEYLHSEIMAHOKFALRLGSMWNKIMSYSDFFQIFQACLRPEPDSE 352
307  LVSTARNNQSRLTSVLHIGAVACQAFQALLGLMFOEFLVRHVLFRKTFSELVQRKQ--E 363
353  NPCLISRLMLWDKILKGARKLHLELIFSSPFMEMEYKLFAMEFVYKYQLOKEVISDD 412
364  TFC-IRHILEYDVWLKMTARTCWRLHLSGMLMEYDNKMILAQEFSRRYATIVEDISD 422
413  HDRSISITALSVOQMTPTLARHLLEEONVISVITETLLEVLPEYLDNRNKNFQO--YS 470
423  HDHAFSIVLSVQVLTTPVPSIAHLLHAHEGIDFKLLHTFYHVAIEKFIIRNKTJLHFSKNIAS 482
471  QDKLGRVYAVICDLKYILISKPTIWIETRLMQFLGFRSFLKILTCMOGMEETIRROVGQH 530
483  LTFEKRANYILYDLRYLLSLAPDVLNDRNGFLEGCRALMRVLNVMQGMESWTRQTGOH 542
531  IEVDPDWEAATAIQMLKNIILMFQEWACACDEELLVAYKECHKAVMRGTSFISSSKTV 590
543  MDYEPWECAFNHLKLTATTSIQVIDWASGDVKLLKLYKMTMRALV--SNSFIVGGEKV 600
591  VQS---CGHSETKSYRVSEDLVSIHLPLSRTLGLHVLRLSLGASVRLHFEVSDFOV 647
601  MQPKKADHVANCLVYDISQVPSIHLPLSRFFAGIYL---HLGA---HD-LTYDGLQT 652
648  EV-----LVVEPLRCLVLVAQVVAEMWRNGLSLISQVYVYODVKCREMYDKD 696
653  ETEALSTKLPRELIEPVLCQTQAMIAQVAGLWRRNGYTLHLQLFYRNVRCRMLDRD 712
697  IIMLQIGASLMDPNKFLLLVLQRYE---LAEAFNKTIISTKD--QDLIKQYNTLIEMLQ 750
713  IACLIQIGASLMSNEFLIHLNRRNTIPWLQENYWSLLSGNEMNDIIRE-ASIDFLE 771
751  VLIYIVGERYPVGVNVTKEVMTREITHLICIEPMPHSAIAKNLPENB--NNEYGLENV 808
772  LLVIIVIGERWMPGVSMVTEEDRLKEIILCLIKPYSHELSEALPDGNSGNSDNVFEV 831
809  INKVATFKKP-GVSGHGVYELKDESLKDFNMVYHYKTOHSAKHMOKRKRQENKDEA 867
832  INTVAFKKPVGADSKGVYELKEHLLEKFNMTFYHTKEDSKAELOERRAKKQLVC 891
868  LPPPPPPFCFPAFSKVINLLNCDIMMYILRTVFERAIDTSDNLWTEGLMQAFHIALGL 927
892  CPPPMLPKLTPTAFTPMANILOCPVFLNCSLIMERALNAYSRSFTESHQKVLHLLGYAI 951
928  LEEKQLOKAPEEVTFDFYHKASRLGSSAMNIOMLEKLGIPOLEGOKDMITWILQMF 987
952  QEE-----LSEHYPFLSFYERSQRYG-----ILEKLEELARCPLEAHYDVLVTIERF 1000
988  DTVKRLREK-----SCLIVATTSGSEIKNDEITHDKEAKERKAEARLHRQK 1037
1001  ---KOLQAKQAPSDGRAPSC-SOQGTGKGLSLSAE---QARENRNRLAERAAH 1052
1038  IMAQMSALQKNFIEETHKLMDYNTSEMPGKE-----DSIMEESTPAVSDYSRIA-L 1087
1053  IMAQMQAKQASFTISSNAEMFADTNETRKESASTGPMDWEDIPPEEQCAVALESKVACL 1112
1088  GPKRGPSTKEVLTCLQCQEEQEVKIENNAVLISACVQKSTALTQHRGKPIELSGEALD 1147
1113  GPKRKFYHGTDDTKFCILCFENCALSIRGGRQLVSSAFVQTSRV----- 1155
1148  PLFMDPDL---AYGTGSCGHVMHVCWQYF---EAVQLSSQRIHVLDLDESSEVL 1201
1156  -IETPTNLNRNSQSAHLSHSCGHVMHYSWLEYFTNEEFKELRRHRNRAALAAQAAVFEQ 1214
1202  CPLCKSLCNTVPIPIPLQPPQ-----KINSENADALAQL-----LTAR 1239
```

```
Db 1215 CPCRTLSNIIPTVETPLAFSPAPPSPNESYLPDSEVMSLAIETLGNVKBHETLTP 1274
QY 1240 WICOTVLARISGYNIRHAKGENPIPIFNQMGDSTLEFHSILGFSVESSIKYSNIKEMV 1299
Db 1275 SVSNIL-RLSGVVGGLAQPERSVOLIKNPP-----RLHADYIEGIE-----FLKKA 1319
QY 1300 ILFATYIYRIGLK-VPPDERDRVPMLTWTCAFTTQAIENLGLDDECKPFLGALONRQHN 1358
Db 1320 LLNTMKIQSHLKHDPFAIESIEMVPL-WDSCSYTTLQALETYLYAVEKPLKAELSMRHQ 1378
QY 1359 GLKALMO-----FAVAQRITCPVOLIKHLVRLSVLPNIKSEDTCLLSIDLFLVLV- 1412
Db 1379 CARNLVACRSRSALEWETDLPLPWRMSQAEFSRLLDIFNQDTSVLEWDCFRVLVP 1438
QY 1413 ---GAV-LAPPSLYWDVDPQLPSSVSSSYNHLILFHLITMAHMLQILLVDTGLPLAQ- 1467
Db 1439 FQFGLNLMVPEKYTIIP-----SGSMFDFVIMOTMFLAQLTKAVLCFQVEKEKAKR 1492
QY 1468 -----VOEDSEE--AHS 1477
Db 1493 AEKAPNSLTQLDYIEQLPSRIRDNMIDFYRRYNIPARVLQTKQKQLVEESEPENQGHG 1552
QY 1478 -----ASSFEAISOYTSSTCCDIPGWYLVWSLKNGTIPYLRCALFFHYLLGVTPP 1530
Db 1553 QTVVIPCESHLLALLEY-----VORQSSFLRCSCLFRFLTDVDFP 1595
QY 1531 EELHTNSAEGEYSALCSYLSLPTNLELLFOEYWDVTRPLL-----QRRCAD 1576
Db 1596 DTFPTDQPD-RFDLMQCYGLDPLMGVYFD--METVYATMHSFASHPHIDREVEQRQCP 1652
QY 1577 PA-----LLNCLQKQNTVYPRKRNLSIELPDDYSCLLQNAQHFRCPSADDERKHPVL 1631
Db 1653 DARRSLQVWPLRP-----LPR-----LKVLCDDFSLLNSVSDIFCPNNEREEMKPTM 1702
QY 1632 CLFCGATLCSQTCQCEVNGEEVGCIFHALHC-----KARGCAY 1672
Db 1703 CLICGLLTCGSCYCCQELQKSVGATHAHAGAEVGLFRLTRDCQVYVLRGKGCYV 1762
QY 1673 PAPYLDYGTDPGLKRGNPLHLSRERYRKLHLWQOHCITEELARSQETNMQLFGFNW 1731
Db 1763 PPPYLDYGTDMGLRGNPLRLSQAYRKIYLQWLGHGLGELTARLNDANVANVAQAQW 1821

RESULT 13
Q9VX91 PRELIMINARY; PRT; 1824 AA.
AC Q9VX91
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE CG9086 protein.
GN CG9086
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Balley R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
```

```
Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischer W.,
RA Foster G., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson K., Nixon K., Nusskern D.R., Paclet J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zeng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RA "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA Paclet J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RA "Sequencing of Drosophila melanogaster genome."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,
RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RA "Annotation of Drosophila melanogaster genome."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE003504; AAF48687.2;
DR FlyBase; FBgn0030809; CG9086.
DR InterPro; IPR003126; znf_Nrecogin.
DR Pfam; PF02207; zf-UBR1; 1.
DR SMART; SM00396; znf-UBR1; 1.
SQ SEQUENCE 1824 AA; 208359 MW; 79992c1117582a0 CRC64;

Query Match 24.2%; Score 2208.5; DB 5; Length 1824;
Best Local Similarity 31.1%; Pred. No. 8.1e-145;
Matches 595; Conservative 300; Mismatches 706; Indels 313; Gaps 55;
```

```
QY 10 EAGGTERMEISAEIPOTPORLASWMDQOVDYFATFLHLHLAQLVPEIYFAEMDPDLKQOE 69
Db 29 QAGTLDORSIIIEFLAKRESPK---YFDYQT---SATVKDNTNITLTKCMFKE---SLAKEE- 78
QY 70 SVQMSIFTPLEWFLGCEPDICLEKLKHSQ-AFQLCGRVFKSGETTYSCRDCAIDPTCVL 128
Db 79 -----IIDVVVEFLGNDPSSALEKLRLBGNATVCGKVFKEGPEYTSRECGVDPTCVL 133
QY 129 CMCDFQDSVHKHNRKYMHTSTGGFCDCGDTBAWTKGPCVNHPEGRAGTIKENSRCPLN 188
Db 134 CVNCFKRSRAHRPHKYKMTSGGGCCDCGDDDAWKDQYCELHLANKKNPLESKI---LT 190
QY 189 BEVIVQARKIPSVIKYVVEWMTIWEEREKLPPELQIREKN-----FRYCYVLENDEH 240
Db 191 DAVLERVEICFAILLAFCSYSL-----ETEPNASLQCLDGNVGGQVDGAQYCTVLYNDES 246
QY 241 HSYDHIYSLORALCELAELAAHLHTAIDKEGRRRAVKAGAYAAQCAKEEDIKSH----- 294
Db 247 HTFDVOIOTLTKIAKRAKADAMEIIVAAIDREGRVAVKCDTFEENKLVSIENQMLPTS 306
QY 295 ---SENVSOPHLVHEVLHSEIMAHQKFAIRLGSWMNKMIMSYSDFRQIFCQACLREEPDSE 352
Db 307 LVSTARNQSLRTSVLHIGAVACQOFALQLLGNFQOEFVRLHFLFKTFSELVQRKQ---E 363
QY 353 NPCLISRLMLDAKLYKGARKILHELIFSPFMEYKKLFAMEFVKYKLOKEIYSD 412
Db 364 TFC-IRHILEYDVKLWKTARTCWRLLSISGMLEMDYDNKMILAQEFSRRYATIVEDFISD 422
QY 413 HDRSISIALSVQMETVPTLARHLIEEQNVISVITETLLEVLPEYLDNRNKNFNQO--YS 470
Db 423 HDHAFISIVLSVQLETVPSTIAHLLAHEGIFDKLLHTFYHVAIERFIRNKTLFHESKNIAS 482
QY 471 QDKLGRVAVICDLKYILISKPTIWTIRLMOFLEGFRSEFLKILTCMQGMEIRROVGQH 530
Db 483 LTFEKRANYILYDLRYLLSKPDVLSLNDLRNGLFEGCRALMRVLNVQMGMESMTROTGOH 542
QY 531 IEVDPDWEAAIAIOMQLNILLMFQEWACADELLLVAYKECHKAVMRCSTFISSSKTV 590
Db 543 MDYEPWECAFNHLHIKATTISQVIDMASGDVKKLRLKYKMTMRALV--SNSFIVGGEV 600
QY 591 VQS---CGHSLETKYRVSEDLVSLHPLSRTLAGLHVRLSLRGAVSRUHEVSEDFQV 647
Db 601 MQPKKVAHDVANCIVYDSVQPSVSLHPLSRPFAGIYL---HLGA---HD-LTYDGLQT 652
QY 648 EV-----LVEYPLRCLVLAQVVAEMWRNGLSLISQVYFYQDVKCEEMYDKD 696
Db 653 ETEALSILTPREIIEPVLCQAMIAQAGLWRNGYTLHLQLYFYRNVRCRVEMLDRD 712
QY 697 IIMLQIGASLMDPNKFLLLVLQRYE---LAEAFNKITSTKD--ODLIKOYNTLIEMLQ 750
Db 713 IACIQIGASLMESEFLHVLNFRNFTIPWLQENYWSLISGNEMNDIIRE-ASIFDEFLE 771
QY 751 VLIYIVGRYYPGVGNVTKEEVTREIITHLLCIEPMPHSAIAKNLPEN--NNETGLENV 808
Db 772 LLVIIGERKMPGVSMVTEEDRLREIIQLLCIKPYSHSELSRALPDGNSGSDNVFEV 831
QY 809 INKVATFKP-GVSHGVYELKDESLEKDFNMFYHYKTSQHSKAEHMQRKRQKQENKDEA 867
Db 832 INTVAVFKKPGVADSKGYELKEHLKKEFNMFYHYTKEDSKAELOQRERRKAKKQLVC 891
QY 868 LPPPPPPFCPAFSKVINLLNCDIMMYTILRTVFERAIDTDSNLWTEGMLQMAFHILALGL 927
Db 892 CPPMPLKLTFAFTPMALIQCPVFLNCSLIMERALNAYSFSFESHQLVHLHLYGAI 951
QY 928 LEEKOOLKAPEEEEYTFDFHKASRLGSSAMMIQMLLEKLKIGIPELEGOKMITWTLOMF 987
Db 952 QEE-----LSEHYFLPFYERSQEYG-----ILEKLEELARCPLEAHYDFVLTIERF 1000
QY 988 DTVKRLRKPK-----SCLIVATTSGSESIKNDIETHDKAEKRAKAEARLHRQK 1037
Db 1001 ---KOLQAKAPSDGRAGPSC--SQQGTGKGLSAAE---QAREERENRARLAAERRAH 1052
```

```
QY 1038 IMAQMSALOKNFIETHKLMYDNTSEMPCKE-----DSIMEESTPAYSDYSRIA-L 1087
Db 1053 IMAQMOQAQKSFSSNAEMFADTENETRKESASTGPMEDIPPEEEOGAVALSKVACL 1112
QY 1088 GPKRGPSTVEKEVLTICILCOEQEVKTENNAWVLSACVQKSTALTQHRGKPTLSGEALD 1147
Db 1113 GPKRKFYHGTDITFKCILCENCAISRGGRLVSSAFVQTSRV----- 1155
QY 1148 PLEWMDPL---AYGTYTGSCHVMHVCWKQYF---EAVQLSSQORIHVDLFDLESGBYL 1201
Db 1156 -IETPTNLRNSQSALHISCCHVMHYSWLEFYFTNEEFKELRRPHRNRALAAQANVEFQ 1214
QY 1202 CPICKSLCUNTPIPIIPLQPO-----KINSENADALAQL-----LTAR 1239
Db 1215 CPTCRTLNSAIIPVTETLPAFSAPPSPNESYLPDSEFVEIMSTLAIELGNVKDHELTLPL 1274
QY 1240 WIOTVLARIISYNIIRHAKGENPIPIFFNOGMDSTLEPHSILSGVESSIKYSNISIKEMV 1299
Db 1275 SVSNIL-RLSGVVGGLAQFERSVQLIKNPP-----RLHADYIEGIE-----FLKKA 1319
QY 1300 ILFATTIYIRIGLK-VPPDERDPRVPMLTWTSTCAFTIOAIENLLDGDGKPLFGALQNROHN 1358
Db 1320 LLNTMKIQOHLKDHDPALIESIEMVPIL-WDSCSYTLOALEIYLYAVEXPKLAELSMRQHS 1378
QY 1359 GLKALMQ-----FAVAQRITCPQVLIQKHLVRLLSVVLPNIKSEBDTPCLLSIDLFHVLV- 1412
Db 1379 CARNLRACRSRSALWETDLPPLPMRQSAEFSSRLDITFNQNDTSVLEWDCFRVLVP 1438
QY 1413 ---GAV-LAPPSLYWDDPVDLPSSVSSYNNHLYFLHLITMAHMLQIILLTVDTLGLPLAQ- 1467
Db 1439 FQFVGLNLMVPEKGYKTIIP-----SGSMFDFYIMQTMFLAQLTKAVLCDFDVEKEKAKR 1492
QY 1468 -----VQEDSEB--AHS 1477
Db 1493 AEKAPNSELTDQDIYEQLPISRIRDNMDIFRRYNIIPARVLOKTOKQOLVEESEENQGHG 1552
QY 1478 -----ASSFAEISQYTSIGSGCIDPGWLYMSLKNGITPYLRCAALFHYLLGVTPP 1530
Db 1553 QTVVIPCESHHLALLEY-----VQRMSSFLRCSCLFYRFLTDVDFP 1595
QY 1531 EELHTNSAEGEYSALCSYLSLPTNLLFQBYWDTVRPLL-----QRRCAD 1576
Db 1596 DTEPTDQPD-RFDLMCOYLGLDPLMGVYFD--METVYATMMHSFASHPHIDREVBQRC-Q 1651
QY 1577 PALLNCLKOKNTVVRPRKNSLIELPDDYSCLLNQASHPRCPBSADDERKHPVLCLFCG 1636
Db 1652 PDARRSLQVPECLRPLPR-----LKVLCDDFSDLINSVSDIFCPNNEREEMKTPMCLICG 1707
QY 1637 AILCSNICCOEITVNGEEVGACIFHALHC-----KARGCAYPAPYL 1677
Db 1708 LILCGOSYCCOPELGVSVGACTHHACGAEGVIFLIRDCQVYVLRGKGCFFPPPYL 1767
QY 1678 DEYGETDPGLKRNPLHLRSERYKRLHVMQOHCIIIEIARSQETNQMFGFNW 1731
Db 1768 DEYGETDMGLRRGNPLRLSQAAYRKVYLQWLGHGLHGEIARLNDNANVAAAAQW 1821
```

## RESULT 14

P91133

ID P91133 PRELIMINARY; PRT; 1927 AA.

AC P91133;

DT 01-MAY-1997 (Tremblrel. 03, Created)

DT 01-MAY-1997 (Tremblrel. 03, Last sequence update)

DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)

DE Hypothetical 218.1 kDa protein.

GN C32E8.11.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;

OC Rhabditidae; Pelodierinae; Caenorhabditis.

OX NCBI\_TaxID=6239;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Bristol N2;

RX MEDLINE=99069613; PubMed=9851916;  
RA None;  
RT "Genome sequence of the nematode C. elegans: a platform for  
RT investigating biology. The C. elegans Sequencing Consortium. #;  
RL Science 282:2012-2018(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RA Gattung S.;  
RT "The sequence of C. elegans cosmid C32B8. #";  
RN Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.  
RL [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RA Waterston R.;  
RT "Direct Submission. #";  
RN Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
RL EMBL; U88308; ABA42328.1;  
DR WormPep; C32B8.11; CE08535.  
DR InterPro; IPR003126; Znf\_Nrecognin.  
DR Pfam; PF02207; zf-UBR1; 1.  
DR SMART; SM00184; RING; 1.  
DR SMART; SM00396; Znf-UBR1; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 1927 AA; 218097 MW; 6C2A581C52A8E9CF CRC64;  
  
Query Match 20.6%; Score 1886.5; DB 5; Length 1927;  
Best Local Similarity 27.3%; Pred. No. 2.6e-122;  
Matches 553; Conservative 305; Mismatches 636; Indels 513; Gaps 67;  
  
QY 102 QLAGRVKSGETVYSCRDCAIDPTCVLCMDFQDSVHNKHYKMHSTGTGGFCDCGDTGA 161  
DB 14 QICGHVEKNGELYITCLDCATDCTVMCLQCFEYIHKSHKYMHSMSGYCYDCGDADA 73  
  
QY 162 WKTGPPFCVNNHPG-----RAGTIKENSRCPLNEEVIVQARKTFFSVIKYVEMTWEE 214  
DB 74 WTEGYACANHEKKDEEAVALPELKK--RCEQLVEIILQ---FSLSMITHKDDKLPEI 128  
  
QY 215 EKELPPELQIREKNERYCYVFNDEHSHYDHYVLSQALDCELAELAQHTAIDKEGR 274  
DB 129 FEKKPKE--VTNEAQQVLTLYNDETHYTSVIVKLYETHICTKDOAMLVATIVDREGS 186  
  
QY 275 AVKAGAAQCEAKEDIKSHS-----ENVSOH--PLHVEVLHSEIMAHOKFALRLGSW 325  
DB 187 AVKLGKADCTKAKDDVORTARDPTSIKSSNNHPLSVKVMYDTTFLALQNFSLTLTW 246  
  
QY 326 MNKIMSYSSDFRQIFCOACL----- 345  
DB 247 LNTQMDVFPPLREIVGEILLSSKFKALKNYTRKMKSEDRQLVAGIIRNVMLPDEEEL 306  
  
QY 346 -----REEP-----DSENPLCI 357  
DB 307 FALDGRMDVEMDDDDDIGEALQMEIDADDEEITAAALAGVSEHQSPGSRDSSTETML 366  
  
QY 358 SRLMLDAKLYKGARKTLHELIFSSFFEMEKYKLFAMEFVKYKQLOKEVYISDDHRSI 417  
DB 367 ENILLQDTQMWKAGRSILHOMLMTVMFYDQKVRFAKAFMLHYNEIYEDFIKDDHMDV 426  
  
QY 418 SITALSVMQMTVPTLARHLLEEQNVISVITETLLEVLPEYLDNRN-----KFNFGYS-Q 471  
DB 427 SVVGLSVQFMTVPSLARKLVAEDQAFSVISKAIRDQTDKPKVYVYNDGKIARFDTSRFP 486  
  
QY 472 DKLGRRVAVICDLKYLILSKP--TIWTERLMQFLEGFRSPLKTLTCMQGMEETRRQVG 529  
DB 487 PELRRSLHTRDMAYILNAPVSESDWNRELIDGVEQGFADFLFLQHLQGMDEVKQAVE 546  
  
QY 530 HIEVDPDWEAIAIQMLKNIILMFQEWACDEEL---LLVAYKECHKAVMRCSTSTISS 586  
DB 547 HQWSESEWAFNTLLRLKDAISMIIIGWAETNEEVHRLMTM---CLELMNRMPVYTKS 603  
  
QY 587 S-----KTVQSCGHSLETKSVYRVEDLVSTHLPLSRTLAGL-----HV 625

DB 604 EEDTYELTVTINGESCRISHFDVLKSSTSVHQPVVRIITAGLFSASNYTGFLNRRSNNSH 663  
QY 626 RLSRLGAVSRLEHFVSFEDFQVEVLEPLCLVLVAQVAVMWRNGLSLISQVFPYOD 685  
DB 664 SLNQ-----ERIKELINCKD--ETNLYELSLRVLVLAQSNATLWRNGFSLNIQHNYFS 717  
  
QY 686 VKCREEMWDKDIIMLOIGASLMDPNKFLLLVLQRYEL-----AAFNKTIISTKQD 735  
DB 718 PLCRNEMFDRDVLMMQVGAALTPTKTFIHLHLRFLFKWATSEFQDRAKANEKPAKPESE 777  
QY 736 DLIQYNTLIEEMLOVLIVIGERYVPGVGNVTKEEVTMRIIHLCLIEPMHPHSAKAKNL 795  
DB 778 DLSKTLVMAIEFIOCLILILICERTYGVGKTRPMDQMKREVIIHLITGTSHTFSHIQQKM 837  
  
QY 796 PENENNE--TGLENVINKVATFKKPGVSGHGYVELKDESLKDFNMFYHYSKTHQKAEHM 854  
DB 838 SHDINSKRLSLHEAVNLVADFRKPLATTAGOFCHKESLSLPTYSPEFMHYKSDQSAAEQS 897  
  
QY 855 QKK--RRKOENKDEALPPPPPEFCFAPFSKVINLLNCDIMMYILRTVFERAIDTDSNLWTE 913  
DB 898 QARVRKMEKSVRACAPPILPDFQTFERIPELITNVLIVHVRLLIDRTA--RRSRFSSD 956  
  
QY 914 GMLQMAFHILALGLEEKQOLQKAPEEVTFDYHKASRLGSSAMNIOMLLEKLGIPOL 973  
DB 957 RLFHKTLYLIGIALNEE-----EKNPSPGFTQRA---BESTGLLSLLEGLVGKP-- 1002  
  
QY 974 EGQKDMITWILQMFDTVKRLREKSLIVATTSGSES--IKNDEITHDKAEKRAKAAEA 1031  
DB 1003 --ESSICILLEV--TVEKYRK--LIKARAGVPEAAPAPENKPAQSEIEIKAKRAARA 1055  
  
QY 1032 RLHRQKIMQAOSLAKQNFETHKLMYD-----NTSEMPG-----KED-----STMEERS 1075  
DB 1056 EM-RQKAMAKMSNMQSKEM--KKTEDEKDKESQTPSEKSEVYVKDDYDNKHPFDEDV 1111  
  
QY 1076 TPVSDYDRIALGP--KRGPSVTEKEVILFCILCOEVEV--KIENNAMVLSACVQKSTALRQ 1133  
DB 1112 VKQVGHDPVPCIGANKHAEVLKPRTLTLCILCOEDEIITAPQOQKPMVCVAAFIQQSOLF-- 1170  
  
QY 1134 HRGKPIEL-----SG--EALDPLFMDPLAYGTYTGTSCGHVMHVAWCWKYFEAVOLSSQOR 1187  
DB 1171 HKNKGELMTASSGISTRDLLTAPATIQYVDVSTCSHMYECTRSIAEAMRSRESUR 1230  
  
QY 1188 IHV-----DLFDLESGEVLCPLKSLCNTVPIIP---LOPKQ-----INSENADAL 1231  
DB 1231 ARQVQCHSHKVDVTENGEOYQCPCLCKRLSNAAPVLPVLPAYOLTQNGFSTVSGAGKENFTD- 1289  
  
QY 1232 AOLTLFARQIVTLARI-----SGYNIRHAKGENPI-----PIFFNQGMG 1271  
DB 1290 -----WVARVKRNLEMPLESSESVSXKKGHSKRSHRSERSLDLLEKLSKDPDTANTSAG 1341  
  
QY 1272 DSTLEFHSILSGV-----ESSIKYSNISKEMVI----- 1300  
DB 1342 -----VLOFGAMEMSSATHMPASBSOMLMTTSPSODDVEFYNELAAMFVDDVNNVT 1393  
  
QY 1301 ---LFATTIVRIG--LKVPDPDERPVRPM----- 1324  
DB 1394 TSPAATPETIPAIGSSSRIPESQESGKKPLSSQIQHVLYSLIRPPALLNSNRCSSPE 1453  
  
QY 1325 -----LTWSTCAFTIOAIENLGLDEG 1345  
DB 1454 GFEEPIKDLGKNMKFRKRGNELKTNFIEKHLKGVYISTVWQSTAHVARAISTSYLHYDK 1513  
  
QY 1346 KPLFCALONRQINGLAL-----MQFAVAQRITCPQVLIQKHLVYR-LLSVVL--- 1391  
DB 1514 KPLFCALATRQRDCLSAMARLCSLHNNQFLL-----HAYSDMLRVLFLCEP 1560  
  
QY 1392 PNKISEDTPCLLSIDLHVLVGAFLAPSLYWDPDVLOPSSV--SSSYNHLVFLHLITMA 1450  
DB 1561 PRPKLAQTP-----GSPLLSAPSTSSFTFA---PAQIPHSGTTFNFAFLVQLFNPA 1606  
  
QY 1451 -----HMLQILLTVDTLGLPLA-----QVQDSESEAHSASSF 1481  
DB 1607 GPRKNVNLILQIDI-----LSLIAITHSEADGNDVNMEEBQESQMEVDPVAAQIRKL 1661



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 1, 2002, 12:39:35 ; Search time 4409.56 Seconds  
(without alignments)  
23599.681 Million cell updates/sec

Title: US-09-724-126A-1  
Perfect score: 6308  
Sequence: 1 gccagaattcgccagcagg.....aattttgtatttggtgtttt 6308

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues  
Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

- 1: gb.ba.\*
- 2: gb.htg.\*
- 3: gb.in.\*
- 4: gb.om.\*
- 5: gb.ov.\*
- 6: gb.pat.\*
- 7: gb.ph.\*
- 8: gb.pl.\*
- 9: gb.pr.\*
- 10: gb.ro.\*
- 11: gb.sts.\*
- 12: gb.sy.\*
- 13: gb.un.\*
- 14: gb.vi.\*
- 15: em.ba.\*
- 16: em.fun.\*
- 17: em.hum.\*
- 18: em.in.\*
- 19: em.om.\*
- 20: em.or.\*
- 21: em.ov.\*
- 22: em.pat.\*
- 23: em.ph.\*
- 24: em.pl.\*
- 25: em.ro.\*
- 26: em.sts.\*
- 27: em.sy.\*
- 28: em.un.\*
- 29: em.vi.\*
- 30: em.htgo\_hum.\*
- 31: em.htgo\_inv.\*
- 32: em.htgo\_rod.\*
- 33: em.htg\_hum.\*
- 34: em.htg\_inv.\*
- 35: em.htg\_rod.\*
- 36: em.htg\_other.\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

8

Result No.	Score	Query Match	Length	DB ID	Description
1	4435.4	70.3	6395	6	AR030784
2	4435.4	70.3	6395	6	AR121463
3	4435.4	70.3	6395	10	AF061555
4	2546.8	40.4	2550	9	AK027803
5	996.2	15.8	1001	6	AK030785
6	996.2	15.8	1001	6	AR121464
7	994.2	15.8	999	9	AF061556
8	981	15.6	6158	9	AB002347
9	764.2	12.1	818	9	HX521
10	651.6	10.3	2958	9	AK026998
11	503.8	8.0	16518	9	AC090514
12	487.8	7.7	2512	9	AK026948
13	354	5.6	190727	9	AC068724
14	279.4	4.4	61896	2	AC090992
15	215.2	3.4	910	10	MM1UBR2
16	212.2	3.4	107304	2	AC016274
17	212.2	3.4	164468	2	AC021899
18	180.8	2.9	164468	2	AC021899
19	160.6	2.5	3980	9	HS1UBR4
20	150.2	2.4	107304	2	AC016274
21	132	2.1	28834	2	AC017982
22	132	2.1	152545	3	AC010922
23	132	2.1	179016	3	AC018489
24	130.4	2.1	900	9	HS1UBR3
25	123.4	2.0	907	10	MM1UBR1
26	115.2	1.8	61896	2	AC090992
27	111.8	1.7	2616	10	MM1UBR6
28	106.4	1.7	830	9	HS1UBR5
29	104	1.6	143291	9	AL137792
30	104	1.6	143697	2	AL358572
31	104	1.6	295265	2	AL353354
32	102.6	1.6	716	10	MM1UBR5
33	101.8	1.6	1000	9	HS1UBR6
34	100.8	1.6	1370	9	HUMORF02
35	100.4	1.6	1100	10	MM1UBR4
36	97	1.5	179588	2	AC009852
37	97	1.5	182660	2	AC012416
38	97	1.5	183982	2	AC009825
39	91.4	1.4	1000	9	HS1UBR7
40	91.2	1.4	727	10	MM1UBR7
41	89	1.4	789	4	CFU12687
42	88	1.4	168639	2	AP001992
43	88	1.4	213613	2	AC013405
44	86.4	1.4	400	9	HS1UBR1
45	86	1.4	195064	2	AL138921

ALIGNMENTS

RESULT 1					
LOCUS	AR030784	6395 bp	DNA		
DEFINITION	Sequence 1 from patent US 5861312.				
ACCESSION	AR030784				
VERSION	AR030784.1	GI:5943998			
KEYWORDS	Unknown.				
SOURCE	Unknown.				
ORGANISM	Unknown.				
REFERENCE	1 (bases 1 to 6395)				
AUTHORS	Varshavsky,A. and Kwon,Y.Tae.				
TITLE	Nucleic acid encoding mammalian UBR1				
JOURNAL	Patent: US 5861312-A 1 19-JAN-1999;				
FEATURES	Location/Qualifiers				
source	1..6395				
BASE COUNT	1802 a 1376 c 1519 g 1698 t				
ORIGIN					

29-SEP-1999

PAT



Query Match		70.38;	Score 4435.4;	DB 6;	Length 6395;
Best Local Similarity		86.33;	Pred. No. 0;		
Matches 4997;		Conservative	0;	Mismatches	721;
				Indels	69;
				Gaps	6;
Qy	590	ttcaggggccgtcgtaaaagtgtcgtccctgtctc-tccgaccgcccacaggtttccgct	648		
Db	8	ttcaggggccgtcgtaaaagtgtcgtccctgtcgtccgtcgccgcccacaggtttccgct	67		
Qy	649	tgcctctggccggggtcgccaactgcaggcgtcagtttccctcgaagtggcgacgagg	708		
Db	68	agctggcgccggggtcgccaactgcggcggttcgtttcccttaagatggcgacgaag	127		
Qy	709	aggttgagggtactgagaggtgaaatcagcgggagttaccocagaccocctcagcgtc	768		
Db	128	agatggagcggcgccagagagtgagcgtcagcccgagcctccctggcccgccgacgccc	187		
Qy	769	tggcatctgtgggtgagtcagcaagtgtgatttttatactgcctttcttgatcatattggcac	828		
Db	188	cggtatctgtgggtgagtcagcaagtgtgatttttatactgcctttcttgatcatattggcac	247		
Qy	829	aattggtccagaaatttactttgtcgtgaaatggaccacagcttggaagcagggagaa	888		
Db	248	aattagtcgacgaattttattttgctgagatggaccagatttggaaagcaagagaga	307		
Qy	889	gtgtacaaatgtcaatatctactccactggyaatgggtactattttggagaagatccagata	948		
Db	308	gtgtacagatgtcaatatctactccactggyaatgggtactattttggagaagatccagata	367		
Qy	949	tttgccttagaagaattgaagcacagtgagacatttcagctttgtgggaggttttcaaaa	1008		
Db	368	tttgccttagaagaattgaagcacagtgagacatttcagctttgtgggaggttttcaaaa	427		
Qy	1009	gtgagagacaacctattcttgagggtatgtgcaattgatccacatgtgactctgta	1068		
Db	428	gtggagaacaacatatttccctgtagggaattgtgcattgtatccaaactgtgtgctgtga	487		
Qy	1069	tggagctgtccagagacagtggttcaataaaatcatcgttacaagatgcatacttctactg	1128		
Db	488	tggagctgtccaaagtagtgttcataaaaaccatcggttacaagatgcatacttctactg	547		
Qy	1129	gaggaggttctgtgactgtgagacacagaggcatggaactggccctttttgtgtaa	1188		
Db	548	gagggggttctgtgactgtgagacacagaggcatggaactggccctttttgtgtgg	607		
Qy	1189	atcagaacctggagagcaggtactataaagagaattcacgctgtcgtttggaatgaag	1248		
Db	608	atcagcagccttggaaagacaggtactataaagagagcttacattggcccatTTGAATGAAG	567		
Qy	1249	aggttaattgtccaaagccaggaaaaattttccttcagtgatataaaatgtcgtagaataga	1308		
Db	668	aggtgatttgcctaagccaggaaattttcccttccgttgatataaaatgataaataga	727		
Qy	1309	ctatatgggaagagaaaaaactgcctcctgaaactcagataaggagaaaaatgaaa	1368		
Db	728	ctatatgggaagagaaaaaactgcctcctgaaactcagataaggagaaaaatgaaa	787		
Qy	1369	gatactattgtctcttttcaatgatgaacaccatttcattatgaaccagtcataacagcc	1428		
Db	788	gatactattgtctcttttcaacagatgagcaccatttcattatgaaccagtcataacagcc	847		
Qy	1429	lacaagaagcttctgactgtgagctcgagagggcccatgttgataccactgcatattgaca	1488		
Db	848	tgcagagagctctagatttgcgagcttgcagagggcagctgacacagctgcccattcgaca	907		
Qy	1489	aagaggggtcgtggggtgttaaagcgggagcttatgtcgtctggcaggaagcaaaaggaag	1548		
Db	908	aagaggggtcgtggggtgttaaagcgggagcttgatgcccactttgccaggaagcaaaaggaag	967		
Qy	1549	atataaagagtcattcaaaaaatgtctctcaacatccacttcattgtagaagtattacact	1608		
Db	968	atataaagagtcactcaaaaaatgtctctcaacatccacttcattgtagaagtattacact	1027		

Qy	1609	cagagattatggctcatcagaaaatttgcgttggcttcttggcttctcagatgaacaaaatta	1668		
Db	1028	ccgtggtttatgcttaccagaaaatttgcgttggcttcttggcttctcagatgaacaaaatta	1087		
Qy	1669	tgagctattcaagtacttaggcagatcttttgcgaagcatcctctagagagaacctg	1728		
Db	1088	tgagctatttcaagtacttaggcagatatttttgcgaagccttccctctgtagaagaacctg	1147		
Qy	1729	actcgagaaatccctgtctcatataagcaggtttaatgcttttggggtgcaaaactttataaag	1788		
Db	1148	gctctgaaaaatccctgtcttataaagcagactaatgcttttggggtgcaaaactttataaag	1207		
Qy	1789	gtgcgctgaagatccctcatgaattgatcttcagcaggtttttttatggagatggaataca	1848		
Db	1208	gtgcgctgaagatccctcatgaattgatcttcagcaggtttttttatggagatggaataca	1267		
Qy	1849	aaaaactcttgccttggaatttgaagtattataaaacactgcagaagaatatatca	1908		
Db	1268	aaaaactcttgccttggaatttgaagtattataaaacactgcagaagaatatatca	1327		
Qy	1909	gtgatgatcatgacagaagtatactctataactgcacttccagttccagatgtttactgttc	1968		
Db	1328	ggcagcaccacgagagagacatctccatracccgctgtccgtgcagatgctcacccgtcc	1387		
Qy	1969	ctactctgctcgacatcttattgaagcagaaatgttctctctgtctacttgaactc	2028		
Db	1388	cgaccttggcccgcatcttattgaagcagagaaatgttatttctgtctatttactgaaacgc	1447		
Qy	2029	tactagaagttttacacctgagctacttgacagagaaataaattcaacttccaggttata	2088		
Db	1448	tgttagaagttttacacctgagctacttgacagagaaataaattcaacttccaggttata	1507		
Qy	2089	gccagagacaaattgggaaggtatgtatgtaataatgtgacctaaatatactctgatca	2148		
Db	1508	gccagagacaaattgggaaggtatgtatgtaataatgtgacctaaatatactctgatca	1567		
Qy	2149	gcaaaccccaatatggacagaagattgaagaatgcagttcccttgaaggttttcgatctt	2208		
Db	1568	gcaaaccccaatatggacagaagattgaagaatgcagttcccttgaaggttttcgatctt	1627		
Qy	2209	ttttgaagatttctacctgtatgcagggaatggaagaataccgaagacaggttgggcaac	2268		
Db	1628	tttctgaagatttctacctgtatgcagggaatggaagaataccgaagacaggttgggcaac	1687		
Qy	2269	acattgaagtggatcctgattgggaggtcccttgctatcacagatgcaattgagaata	2328		
Db	1688	acattgaagtggatcctgattgggaggtcccttgctatcacagatgcaattgagaata	1747		
Qy	2329	ttttactcatgttcccaagagtggtgtgtgtgatgaagaacttctacttgtggcttata	2388		
Db	1748	ttttactcatgttcccaagagtggtgtgtgtgatgaagaacttctacttgtggcttata	1807		
Qy	2389	aagaaatgcacaaagctgtgatgagtgagtcaccagtttcatatctagtagcaagacag	2448		
Db	1808	aagaaatgcacaaagctgtgatgagtgagtcaccagtttcatatctagtagcaagacag	1867		
Qy	2449	tagtaacactgtgtgacatagtttggaaacaaagtcctcacagagtatcttgagagatctg	2508		
Db	1868	tagtgcgaatttgcgggtcatagtcggaacacaaatcctacaaagctgtctgagagacctg	1927		
Qy	2509	taagcatatcttgcacctcttaggaccttgcgtgtcttctatgtacgttttaagcagggc	2568		
Db	1928	taagcatatcttgcacctcttaggaccttgcgtgtcttctatgtacgttttaagcagggc	1987		
Qy	2569	tgggtgctgttccaaagactcatgaatttgcgtcttcttggaggaccttcaagtagaggtac	2628		
Db	1988	taggtgctatttccaaagactcatgaatttgcgtcttcttggaggaccttcaagtagaggtac	2047		
Qy	2629	taqtggaatatccttcaagttgt	2688		
Db	2048	tggtgagatgacccgctgcgt	2107		
Qy	2689	gaagaaatggagctgtctcttattagccaggtgttttattaccagaagtgttaagtgcagag	2748		

[illegible]

Db	3188	AAAGCTGAGGCGCTAGGCTTTTCATGCGCAGAAAGATCATGGCCAGCATGCTCTGCCTTACAGA	3247
QY	3820	aaaacttcattgaacctcataaactcatgtatgacaatacatcagaataatgccctgggaaag	3879
Db	3248	AAAACITTCATTGAAACCCACAACACTCATGTATGATATAGTACAGTAGTAAACAGGGAAGG	3307
QY	3880	aagattccattatggagggaagagacaccccgacgtcagtcactactcagtaattgtcctt	3939
Db	3308	AAGACTTCCATTATGAGGAGAGAGACGACCTTCAGCAGTCTAGTGAGGCGCTTAGAATTGCTC	3367
QY	3940	tgggttcctaaacggggtccatctgttactgaaagaaggtgctgacatgcatactcttggcc	3999
Db	3368	TGGGCCCTTAACCGGGGCCCGGCTTTTACCGAAAGAGAGGTGCTACGCTGCATCTCTGCCC	3427
QY	4000	agaagaacagggaggtgaaaaatagaataatgccaatggtattatcgccgtgtgtccaga	4059
Db	3428	AAGAAGAACAAAGAGGTAAAACTAGAAAAATAATGCCATGGTATTCTCAGCATGTGTGCAGA	3487
QY	4060	aatctactgccttaaccccgacagcaggggaaaaacccatagaaactctcagggagaagccctag	4119
Db	3488	AATCCACCGCCCTTAACCCACGACAGAGGGAAGGCTCTGGACCCTTATAGGGAACACTGG	3547
QY	4120	accacttttcagttccagacttgcataatggaacttatcagaagaagctgtggtcatg	4179
Db	3548	ACCCTCTTTTCATGGATCCAGACTTGGCATGGAACTTATACAGGAAGCTGTGGTCATG	3607
QY	4180	taatgcacgcagtgctgtggcagaagtattttgaagctgtacagctgagctctcagcagc	4239
Db	3608	TAATGCATGCAGTGTCTGGCAGAAGTATTTTGAAGCTGTGCAGCTGCAGCTGCAGCAGC	3667
QY	4240	gcattcatgtgacctttttgacttggaaagtggagaatatctttgccccctctttgcaaat	4299
Db	3668	GCATTCACGCTAGACCTGTTTGACTCGGAGAGCGCGAGTAGTACCTATGCCGCTCTGCAAGT	3727
QY	4300	cctctgtgaataactgtgatcccccatattctcttgaacctcaaaagataaaaacagtgaga	4359
Db	3728	CTCTCTGCAACACTGTCAATCCCATATCCCTTTTGACCGCGCAAGAGATCAACAGTGA	3787
QY	4360	atgcagatgctctgctcaacttttgacctggcaggtggatgatacagactgtttctggcca	4419
Db	3788	ATGCGGAGGCTCTGTCTCAACTTTTCACTTTGCCCCGGTGGATACAGACTGTCTTGGCCA	3847
QY	4420	gaatcacggttataataaagacatgctaaaaggagaaaaacc---caattcctctatttct	4476
Db	3848	GAATATCGGTTATAATAATAAAGCATGCTAAAGGAGAAAGCCCAAGCAGTTCCTGTCTTGT	3907
QY	4477	ttaatacaagaatggagagattctactttggaggtccattccactcctgagtttggcgctg	4536
Db	3908	TTAATCAAGGAATGGGGGATTCAACTTTTGAGTTTTTCAATCCATCCCTGAGTTTTGGAGTTC	3967
QY	4537	agcttcctgatataaattcaaatagcatacgaagaaatggttatctcttttggccacaacaa	4596
Db	3968	AGCTTTCGGTGAAATATTCAATAGTATCAAGGAATGGTCAATCTCTTCGCCACAACAA	4027
QY	4597	tttatagaattggattgaaagtgccacctgatgaaagggaatcctcagtcagtcctccatgctga	4656
Db	4028	TTTACAGAAATGGCCTGAAAGTGGCTTCTGATGAACCTAGACCCACGAGTGCCCATGATGA	4087
QY	4657	cctggagacactgcgtcttcaactatccaggcaattgaaaaatctatttgggagatgaagaaa	4716
Db	4088	CCTGGAGCAGCTGTGCGTTTCAACATCCAGGCAATCAAAAACCTGTTGGGAGATGAAGGAA	4147
QY	4717	aacctctgtttgagcactcaaaaataggcagcataatggtctgaaagcattaatgcagt	4776
Db	4148	AACTTCTATTGGAGGCATTTCAAAATAGACAGCATAGCGGTCTTGAAGGCCCTTAATGCAGT	4207
QY	4777	ttgcagttgcacagaggtattacctgtccctcaggtcctgtatcacagaacaatccttggttcgtc	4836
Db	4208	TTGCAGTTGCACAGAGGGCTACCTGCCCCTCAGGTCCTGATACACAAACACTCTGGCTCGGC	4267
QY	4837	ttctatcagttgtcttcctaacaataaaatcagaagaataacacctgcctctctgtctatatag	4896
Db	4268	TCCTCTCAGTTATCTTCTTAACCTCAATCAGAAATACACACGAGGCTTCTGCTCTGTGG	4327

```
QY 4897 atctgtttcatgttttgggtggtgctgtgttagcattcccaatccctgttattggtgatgacc 4956
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4328 ATCTCTTCCATGTTCTGTGGTGGCGCAGCTTTAGCGTTCCCATCTCTTGTATTGGGATGACA 4387
QY 4957 ctgtgtatctcagcctcttcagttagttcttctataaacacatttattcttccatt 5016
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4388 CCGTGGATCTCGACCGCTGCGCACCTTAGTTCTTCATATAACCACTCTATCTCTTCCATC 4447
QY 5017 tgatcacattgacacatgcttcagatactacttactacagtagaca- -----cag 5064
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4448 TGATCACATGGCGCACATGCTTCAGATACCTCTTACACAGATACAGATCTGTCTCCAG 4507
QY 5065 gcttaccccttgcctcaggttccaagaagacagtggaaggtccattccgcattcttcttct 5124
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4508 GGCGCGCGCTTGCCTGAGGGTGAAGAGGATAGTGAGGAGCGCTCGCTGTCATCTGCTTCT 4567
QY 5125 ttgcagaaattctcaatatatacaagtggtccattgggtggtgataattccctggctggtatt 5184
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4568 TTGTGGAAAGTCTCGACACACAGACGCGCTCACTGGGTGGGTGCTCCCGCTGGTACC 4627
QY 5185 tgtgggtctcactgaagaatgcaccccttcttctctgctgctgctgctgctgctgcttctcc 5244
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4628 TGTGGCTCTCCTGAGGAACGGATCACCCCTTACCTCGCTGCTGCTGCTGCTGCTTTTCC 4687
QY 5245 actatttacttgggttaactccgctgaggaactgcataccaaattcttcgagaagagagt 5304
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4688 ACTATTACTTGGAGTAGCTCCGCTGAAGAACTGTTTGGCCAAATCTCTGCTGAAGGAGAT 4747
QY 5305 acagtgaactctgtagctatctatttctacatacaaaatttctctgctcttccaggaat 5364
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4748 TCAGTGCACCTGTAGCTATCTATCTTTACCCACAAATTTGCTGCTTTTCCAGGAAT 4807
QY 5365 attggatactgaagcccttctccagagtggtgctgctgctgctgctgctgctgctgctgct 5424
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4808 ATTGGATACCATAAGGCCCTTACTACAGAGGTGGTGTGGAGATCCTGCCCTTACTCAAGT 4867
QY 5425 gtttgaagcaaaaaaacccctggttcaggtacccttagaagaaataagttttagatagac 5484
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4868 CTTTGAAGCAGAAAGTCTGTGTGTCAGTACCTAGAGAAAGAAATAGTTGTATAGAGC 4927
QY 5485 ttctgtatgactatagctgctctctgaatcaagcttctcatttccaggtgcccaggtctg 5544
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4928 TTCCTGAGGACTACAGCTGTCTCTAAATCAGGCTTCTCACTTTAGGTGTCCACGGTCTG 4987
QY 5545 cagatgatgacgaagacatcgtctctctgctcttctctggtgggctatactattctctc 5604
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4988 CAGATGATGAGCGAAAGCATCTGTCTCTGTCTTCTGTGGGCCCATCTGTGTCTCTC 5047
QY 5605 agaatttctgcccagaaattgtgaacgggaaggttggagcttgcatctttttcag 5664
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5048 AGACATCTGTGGCAAGAAATAGTGAATGGGGAAGAGGTGGAGCGTGGCTTTTTCATG 5107
QY 5665 cacttcacttggagccggagctctgcatcttctcctaaatacagaaatgcccagtggtccc 5724
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5108 CGCTTCATTTGGTGTCTGGAGTCTGCATTTCTTAAATCCGAAATCCGAGGTGGTCC 5167
QY 5725 tggttgaaggttaaagccagagctgtgctatccagctccttacttggatgaatatggag 5784
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5168 TGGTGGAAAGAAAGCCAGAGCTGTGCTCTACCCAGCCCTTACTTGGATGAATATGGAG 5227
QY 5785 aaacagaccttggctgaagaggggcaaccccttcttattctctgtagcggtatcgga 5844
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5228 AAACAGACCCAGGGCTAAAGAGAGAAACCCACTTCATTTATCTCGGAGCGGTATCGGA 5287
QY 5845 agctccatttggctggaacaacactgcattatagaagagattgtcaggagccaagaga 5904
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5288 AGCTGCATTTGGTCTGGCAACAGCACTGCATTTATAGAGAGATGCTCGGAGCCAGGAGA 5347
QY 5905 ctaatcagatgtatttggatttcaactggcaggttactgtgagctccaactctgctcctcag 5964
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5348 CTAATCAGATGCTATTGGATTAACTGGCAGTTACTCTGAGCTTCACTGCTTCTGCTCAG 5407
```

```
QY 5965 acaatcacaaaatgacgacagtagtaaaagctgattcaaaaattatgaaaaactttctgagg 6024
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5408 ACAATCATGAGTGACATCAATAAAGACTGATCTAAATTTCTAGAGAACTTTCTGAGG 5467
QY 6025 gctggaaaagtagttgaggggtcttttgcctatgctccaggttccacttacatcaataaaat 6084
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5468 ACGGGGAAGTATTGGAGGGTCTTTTGATCCATGTCCAGATTCCACACACATTAAATAAT 5527
QY 6085 atttcttaatgagtagtcttcttcaattagcaaacatatgcttcacaggaaaaa-aggac 6143
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5528 ATTCTCTTAATGAATATTGCTTTCAATTATCAAAACATAAGCTTCAAGGGGAAAAACAAGAC 5587
QY 6144 atagatcaaatctgt----- 6160
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5588 ATAGATTAATGTTTATGTTTCTAGAACACATAAAGAAATGCTTTCATCCCAAGTCTAT 5647
QY 6161 atgtctatgatttccaggaatttattcccttcaaatgttctcattctatttattt 6220
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5648 TTCTGCTAATAATTCCAGAAAACTCCTTTCCCTTCATAACTGTCTTAGTTTCATTTTCATAT 5707
QY 6221 catcacttggtagtagaagtcacgtcaaacagttgtagacatttattgtgtggttaac 6280
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5708 CACCACCTGGTTAAGAGGTACATTAAGCATTTGTGGACATTTCTCCATCTGGCTAAC 5767
QY 6281 tcttctgcaatttgtatttgggttt 6307
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5768 ATCTCTGCACCTTTGTATTGTTGGTGT 5794

RESULT 2
AR121463 6395 bp DNA PAT 16-MAY-2001
LOCUS Sequence 1 from patent US 6159732.
DEFINITION
ACCESSION AR121463
VERSION AR121463.1 GI:14105039
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 6395)
AUTHORS Varshavsky,A. and Kwon,Y.Tae.
TITLE Nucleic acid encoding mammalian Ubr1
JOURNAL Patent: US 6159732-A 1 12-DEC-2000;
FEATURES
Location/Qualifiers
Source 1..6395
BASE COUNT 1802 a 1376 c 1519 g 1698 t
ORIGIN

Query Match 70.3%; Score 4435.4; DB 6; Length 6395;
Best Local Similarity 86.3%; Pred. NO. 0;
Matches 4997; Conservative 0; Mismatches 721; Indels 69; Gaps 6;

QY 590 ttccaggccgcgtcgtaaaagtgctgccccgtctc-tccgaccgcccacaggtttccgct 648
Db 8 TTCAGGGCCGCTGCTAAAAGTGTCGCCCTGTCGGCGCCGCCACAGGTTTCCGCT 67
QY 649 tgcctctgcccggggtcggcaactgcagcgtcagtttccctcaagatgcccgaagag 708
Db AGCTGCGCGCCGGGGCTCGGAACTCGGGCGTTGCTTTCCCTTAAGATGGCGACGAAG 127
QY 709 aggttgaggtagtagagatggaatacagcgcgaggttaacccagacccctcagcgtc 768
Db 128 AGATGACGCGCGGAGAGGATGGAGTGCAGCCGGAGCCTCCCTGCGCCGACGCGC 187
QY 769 tgcacattggtggatcagcaagttgattttattatgcttcttctgcatcattggcac 828
Db 188 CGCATCTGTTGGGATCAGAAAGTTGATTTCTATCTGCTTCTTACATCATTTGGCAC 247
QY 829 aattggtgcagaaatttactttgtgaaatggaccacagacttggaaaaagcaggagaaa 888
Db 248 AATTAGTGCAGAAATTTATTTTGTGCTGAGATGGACCCAGATTTTGGAAAAACAAGAAGA 307
```

QY	889	gtgtacaaatgtcaatattcaattccactccactggaatgggtactatttggagagaagatccagata	948
DB	308	gtgtacagatgtcaatattcaattccactccactggaatgggtactatttggagagaagatcccgata	367
QY	949	tttgcttagagaaattgaagcacagtggaacatttcagctttgtggagaggttttcaaaa	1008
DB	368	tttgctttagagaaattgaagcacagtggaacatttcagctttgtggagaggttttcaaaa	427
QY	1009	gtggagagacaacctattcttgcaaggattgtgcaattgtaccacaatgtgtactctgta	1068
DB	428	gtggagaaaaaacatatattctgtgaggattgtgcaattgtaccacaatgtgtactctgta	487
QY	1069	tggactgttccaggacagtggttcataaaaaatcatcgttacaagatgcatacttctactg	1128
DB	488	tggactgtgtccaaaagtgtgttcataaaaaacatcgttacaagatgcatacttctactg	547
QY	1129	ggagagggttctgtgactgtgagacacagagctggaaaactgccccttttgggttaa	1188
DB	548	gagggggcttctgtgactgtggagacacagagctggaaaactgccccttttgggttg	607
QY	1189	atcgtgaacctggaagcaggttactataaaagagaattcacgctgcctgttgaatgaag	1248
DB	608	atcacgagcctggaagcaggttactataaaagagaattcacgctgcctgttgaatgaag	667
QY	1249	aggttaattgccaaagccaggaataatttcttcagtgataaaaatgttcgtgagaatga	1308
DB	668	aggtgatttgcctcaagccagagaaattatcccttgcgtgatataaatacatattgtagaatga	727
QY	1309	ctatatgggaagagaaaaaagactccctcctgaactccagataaaggagaaaaaataaaa	1368
DB	728	ctatatgggaagagaaaaaagaaatttcccttccctgaactgcagataaaggagaaaaaataaaa	787
QY	1369	gatactattgtccttttccaatgataaacacattcatatgaccacgtcatatcacagcc	1428
DB	788	gatactattgtccttttccaagatgagaccatttcgttatgatcatgtgatctacagtc	847
QY	1429	tacaagagccttgactgtgactgcagagcccgagttgataccactgcaattgaca	1488
DB	848	tcacagagccttcagatttcgagcttgcagagccagctgcacagctgcacagctgcacag	907
QY	1489	aagagggctgtcgggctgttaaagcggggagcttatgctgcttccaggaagcaaaagaa	1548
DB	908	aagagggctgtcgggctgtcaaacagaggtgtgtatgccacttgcacaggaacaaaggag	967
QY	1549	ataataaagatcattcaagaaaatgtctctcaaatccacttcatttagaagtatataact	1608
DB	968	atataaagagctcactcagagaaagcttctcagacacccctccactgtggaagtctgcact	1027
QY	1609	cagagattatgctcatcagaattgcttgcgtcttgcttcgttcgttcgttcgttcgttcgt	1668
DB	1028	ccgtggttatggctcaccagaaattcgcttgcgccttgcctcgatgaaacaaatata	1087
QY	1669	tgagctattcaagtgactttagcagatcttttggcaagcatgcctttagagaagaacctg	1728
DB	1088	tcagctatttcaagtgactttagacagatatatttgcacagccctccactgtggaagtctgcact	1147
QY	1729	actcggagaatccctctgtctcataagcaggttaagtgtttggagtgcaaaagctttataag	1788
DB	1148	gctctgaaaaatccctgtctttataaagcagactaatgtcttgggattgcaaaaactttataaag	1207
QY	1789	gtgcccgtaaagatccttcagattgacttcagcagtttttttatggagatggaataca	1848
DB	1208	gtgcccgtaaagatccttcagattgacttcagcagtttttttatggagatggaataca	1267
QY	1849	aaaaactcttgcattggaattgtgaagtattataaacaactgcagagaagaatatca	1908
DB	1268	aaaaactcttgcattggaattgtgaagtattataaacaactgcagagaagaatatca	1327
QY	1909	gtgtgataatgacagaagatactctataactgcacttccagttcagatgtttactgttc	1968
DB	1328	ggcagcaccacagagaagacatctccataacgcgcctgtccgtgcagatgtctcaccgtcc	1387

QY	1969	ctactcgtcgctgcacatcttattgaagagcagaatgatatctctgtctaatctactgaaacct	2028
Db	1388	CGACCTTGGCCGCGCATCTATTGAAGAGCAGAAATGTATTTCCTGTCTATTACTGAAACGC	1447
QY	2029	tqctaagttttaccttgagtaactggacaggacaataaattcaacttcagggttata	2088
Db	1448	TGCTAGAAGTTTACCCTGTAATACTTGACAGAGAACAAATAAATTTCAACTTCCAGGGTTATA	1507
QY	2089	gccaggacaaaattgggaagagtatatgcagtaataatgtgacctaaaagtatatcctcgatca	2148
Db	1508	GCCAGCACAACTGGGAGAGTCTACGCAGTTATATGTGACCTTAAGTAGATATATCTGTGATTA	1567
QY	2149	gcaaccaccaatatatggcacagaagattaaagaatgcagttcccttgaagggttttcgatcttt	2208
Db	1568	GCAAGCCTGTCATATGGACAGAACGATTAAGAGCGCAGTTCCCTGGAAGGTTTCCGGTCTT	1627
QY	2209	ttttgaagattctaacctgtatgcagggaatggaagaaaaatcccgaagacagagtttgggcaac	2268
Db	1628	TTCTGAAGATTCTTACTGTATCAGCGGAATGGAAGAAAATTCAGAAGACAAGTTGGACACAC	1687
QY	2269	acaattaagtgatcctgattggagagctgccattgctatcacagatgcaattgaagaata	2328
Db	1688	ACATTGAAGTGGACCTGACTGGAGGCTGCCATCGCTATACAGATGCCAACTTAAGAATA	1747
QY	2329	ttttactcatgtttccaagagtggtgcttgtgtgatgaagaacctcttacttgtgctttata	2388
Db	1748	TTTTTGCTCATGTTCCAAGAGTGGTGCTGTGTGATGAAGATCTCTTACTGGTGCCTTTATA	1807
QY	2389	aggaatgtcacaaaagtgtagtaggtgtagtaccagtttcatatcttagtagcaagacag	2448
Db	1808	AAGAATGTCCAAAGCTGTAAATGAGGTGGAGTAGTAGTACAAAATTTTCATGTCCAGTACCAAAGACAG	1867
QY	2449	tagtacaatcggtgacatagtttggaaacaagaacctcacagagtatctgaagatcttg	2508
Db	1868	TAGTGCAANTGTGCGGTCTATAGTCTGGAACCAAAATCTCTACAAAGTGCTCTGAGGACCTTG	1927
QY	2509	taagcatacatctgcoactctctaggacacctgtcgtgctctcatgtacgttttaagcaggc	2568
Db	1928	TAAGCATACACCTGCCACTCTCTAGAACACTTCTGCTGGTCTTCATGTACGTTTAAAGCAGAC	1987
QY	2569	tgggtgctgtttcaagactgcatagaatttgctctttttagggactttcaaagttagaggtac	2628
Db	1988	TAGGTGCTATTTCACAGACTGCATGAATTTGTGCTTTTTCAGACGCTTTCAAAGTAGAGGTCC	2047
QY	2629	tagtggaatactctttcacgttgctgagttgtgtgccagagttgtgtcgtgagatgtggc	2688
Db	2048	TGGTGGAGTAGCCCGTGCCTGCCCTGGTCTCTGGTGGCTCAGGTTGTTGCTGAGATGTGGC	2107
QY	2689	gaagaaatggactgtctcttattagccaggtgtttattaccagaatgtttaagtgcagag	2748
Db	2108	GAAGAAACGGGTCTCTACTCATCAGCCAGGTTTTTCTATTATCAAGATGTTAAATGCAGGG	2167
QY	2749	aagaaatgtatgataaagataatcatgcttcaagattggtagatottttaatggatcccc	2808
Db	2168	AGGAAATGTACGATAAAGATATCATCATGCTTCAGATTTGGAGCATCTATTAATGGATGCCCA	2227
QY	2809	alaagttctgttactgttacttcagagfatgaacctgccgaggtttttaacaagacca	2868
Db	2228	ACAAGTCTTGTTACTTGGTACTTCAGAGATATGAACCTTACTGTATGCTTTTAAACAAGACCA	2287
QY	2869	tatctacaaaagaccaggtattgtattaaacaaataataacactaatagaagaatgcttc	2928
Db	2288	TATCCACAAAACCCAGGATTTGATTTAAACAGTATTAATACATTAATAGAAGAAATGCTTC	2347
QY	2929	aggttctcatctatatgttgggttgagcgttatgtacctggagtgggaaatgtagccaaag	2988
Db	2348	AGGTCTCTATCTATATTGTGGGAAGACGTTATGTACTCGAGTGGGAAATGTTTACCAGAG	2407
QY	2989	asagaggtcacaaatgagaaaaatcattcacttgccttgcattgaaccatgccacacagt	3048
Db	2408	AGGAGGTTATAATGAGAGAGATTACTTACTTTGCAATTTGAGCCCATGCCACACAGTG	2467
QY	3049	ccattgccaaaaatttaccctgagaatgaaaaataatgaaactggcttagagaatgctcataa	3108





YNHLPLFLHITMAHMLIILLTTDTDLSPGPPLEAGEEDSEBARCASAFFVEVSOHTDG  
LTCCGAPCWILWLSLNGIITPYLRCAALLFHYLLGVAPPPELFANSAGBESFALSUSYL  
SLPTNULFLQEXYWDITIRPLLQRMCGDPALLKQSAVVYRPRKRNLSLELEPEDYS  
CLINQASHFCRPSADDERKHPVLFCFGAILCSQNTCCQEIIVNGEVBGACVFFHALHC  
GAGVCIPLKIRECRVILVEKGARGCAPYPLDYEYGTDPGLKRGNDLHLSRERYRKL  
HLVWQHCIIIEIARSQETWQMLFGFNQLL"  
5389..6395  
/gene="Ubr1"  
5429..5434  
/gene="Ubr1"

BASE COUNT 1802 a 1376 c 1519 g 1698 t  
ORIGIN

Query Match		70.3%;	Score 4435.4;	DB 10;	Length 6395;
Best Local Similarity		86.3%;	Pred. No. 0;		
Matches 4997;		Conservative 0;	Mismatches 721;	Indels 69;	Gaps 6;
Qy	590	tccsggggcccgtgtaaaagtctgcctctgtctc- 	tcgcagccggccacaggtttccgct	648	
Db	8	TTcAGGGGCCGCTGAAAAGTGTGCTCCGTGCGCTGCGGCGGGCCGCCACAGGTTTCCGCT	67		
Qy	649	tgcctctggcggggtcgcaactcagcgctcagtttccctcaagatgcgcgacgag	708		
Db	68	AGCTGGCGGGCGGGGTGGGAAGTGGCGGCGTTCGTTTCCCTTAAGATGGCGGACGAAG	127		
Qy	709	aggctggagggtactgagaggatggaatcagcgqgaggttaccccgagccctcagctc	768		
Db	128	AGATGGAGCGGCCGAGAGGATGGACGTGACGCCGAGCCCTCCCTGCCCGCCGACGGC	187		
Qy	769	tggcacttgggtgggategcagcaagtgtgattttttatctactgcttcttgcacatttggcac	828		
Db	188	CGGCATCTGGTGGGATCAGCAAGTTGATTCTTACTACTGCTTCTTACATCATTTGGCAC	247		
Qy	829	aattgggtcgagaatttactttctgaaatggaccacagacttgaaagcagaggaaa	888		
Db	248	AATTAGTGCCAGAAATTTATTTTGTGTAGATGGACCCAGATTTGGAAAGCAAGAGAGA	307		
Qy	889	gtgtacaaatgtcaattctcactcactcagtggaatggttacttatttgagaagatccagata	948		
Db	308	GTGTACAGATGTCAATACTACTCTTTTGGAGTGGTACTTATTGGAGAGGATCCGGATA	367		
Qy	949	tttcttagagaaattgaagcacagtgagcatttccagctttgtgggggggttttcaaaa	1008		
Db	368	TTTGCTTAGAGAAATTAACACACAGTCAGCGCTCCAGCTGTGTGGGAAGGTTTTCAAAA	427		
Qy	1009	gtgagagacaacctattcttgcagggtattgtgaattgatcccaacatgtgtactctgta	1068		
Db	428	GTGGAGAAACACATATTTCCTGTAGGGATTGTGCAATTTGATCCCAACGTGTGTGCTGTA	487		
Qy	1069	tggactgctccagagacagtggttcataaaatcactgttacaagatgcatacttctactg	1128		
Db	488	TGGACTGTTCCAAAGTAGTGTTCATAAAACCATCGTTACAAAGATGCATCTTACTGTG	547		
Qy	1129	gagaggggttctgtactgtgtgagacagaggcattggaactggccctttttgtgttaa	1188		
Db	548	GAGGGGGCTTCTGTACTGTGTGAGACACAGAAAGCGTGGAAAACTGGCCCTTTTGTGTGG	607		
Qy	1189	atcatgaacctggaagagacaggtactataaaagagaattcaacgtgtccgttgtaatgaag	1248		
Db	608	ATCAGGACCCCTGGAAGACAGGTACTACAAAAGAGAGGCTTACATTTGCCCATTTGAATGAAG	667		
Qy	1249	aggttaattgtccaagccaggaaaatttcttctcagtgataaaatgtcgtgagaaatga	1308		
Db	668	AGGTGATTGTCTCAACCCAGGAGAAATTTCCCTTCGGTGATGATAAAATACATTTGAGAAATGA	727		
Qy	1309	ctatatgggaagagagaaaactgcctcctgaactccagataaaggagagaaaaatgaaa	1368		
Db	728	CTATATGGGAAGAGAAAAGGAATTTCCCTCTCTGAACTGCAGATAAGGGAGAAAATGAAC	787		
Qy	1369	gatactattgtctcttctcaatgatgaagaccatctcatatgaccacgtcatatacagcc	1428		
Db	788	GATACTATTGTGCTTTTCAACGATGAGACCAATTCGTATGATGTGATCTACAGTC	847		

Qy	1429	tacaaagaagctcttgaactgtgagctgcagagagccagttgcatataccactgcccattgaca	1488		
Db	848	TGCAGAGAGCTCTAGATTTCGAGCTTTCGAGAGCACAGCTGTCACAGACTGCCATCGACA	907		
Qy	1489	aagaggggtcgtcgggctgttaaacgaggagcttactctctccaggaagcaagaag	1548		
Db	908	AAGAGGGTCGCCGGGTGTCAAAAGCAGGTTGTATGCCACTTCCAGGAAGCAAGAGG	967		
Qy	1549	atataaagagtcattcagaaaaatgtctctcaacatccacttcattgtagaagtattacact	1608		
Db	968	ATATAAAGAGTCACTCAGAGAACGCTCTCTCAGCACCCCCCTCCATGTGGAAGTCTGCAC	1027		
Qy	1609	cagagattatgggtcctcatcagaaaatttgcgttgcgttcttgcgttgcgtgatacaaaatta	1668		
Db	1028	CCGTGGTTATGGCTACCAGAAATTCGCTCTGCGCTTGGCTCTGGATCAACAAATAA	1087		
Qy	1669	tgagctattcaagtgacttaggcagatcttttgcacaagcatgccttagagaagaacctg	1728		
Db	1088	TGAGCTATTTCAAGTGACTTTTAGACAGATATTTTGGCAGGCTCGCTCGTAGAAGAAC	1147		
Qy	1729	actcggagaatccctgtctcataagcaggttaatgttttgggatgcaaaactttataaag	1788		
Db	1148	GCCTGAAAATCCCTGCTTATTAAGCAGACTAATGCTTTGGGATGCAAACTTTATAAG	1207		
Qy	1789	gtgcccgtaagatccctcatgaattgatcttcagcaggtttttttatggagatggaataca	1848		
Db	1208	GTGCCGTAAAGTCCCTTCATGAATGATCTTTAGTAGTTTTTTTATGGAGATGGAATACA	1267		
Qy	1849	aaaaactcttgccttggaatttggtaagtattataaaaaactgcagaaagaataatca	1908		
Db	1268	AAAAACTCTTTGCTATGGAATTTGTGAAGTATTAATAAACACTGCAAGAAAGATACATCA	1327		
Qy	1909	gtgatgatcatgacagaaagtatctataactgcactttcagttcagatgtttactgttc	1968		
Db	1328	GGAGACACACAGAGAACATCTCCATTAACCCCTCTGCTCGTGACAGTCTCACCGTCC	1387		
Qy	1969	ctactgtcgtcgacatcttattgaagagcagaaatgttactctctgtcattactgaaact	2028		
Db	1388	CGAGCTTGGCCCGGCATCTTATTGAAGAGCAGAAATGTTATTCTGTCTATTACTGAAACGC	1447		
Qy	2029	tgtcagaagtatttacctgagcttgcagaggaacaaataaactcaactccaggggtata	2088		
Db	1448	TGCTAGAAGTTTTTACCTGTAATACTTGGACAGAACATAATAATTTCACTTCCAGGGTTATA	1507		
Qy	2089	gccagacaaaattggagagatataatgcagtaataatgtgacctaaagtatatctctgata	2148		
Db	1508	GCCAGGACAAACTGGGAAGAGCTACGCAGTTTATGTGACCTTAAGATATATCTCTGATTA	1567		
Qy	2149	gcaaaccccaaatatggacagaaagattgaagaatgcagttccttgaagggttttcgactctt	2208		
Db	1568	GCAAGCTCTGTCAATATGGACAGAACGATTAAGAGCGCAGTTTCTTGGAAAGTTTCCGGTCTT	1627		
Qy	2209	ttttgaagattcttacctgatagcaggggaatgaagaaaatccgagagacaggttgggcaac	2268		
Db	1628	TTCTGAAGATTCTTACCTGTATGCAGGGAATGGAAGAAATCAGAAGAACAAAGTTGGACAAC	1687		
Qy	2269	acattgaagtggatcctgattgggaggtgctccattgctatacacagatgcaattgaaagata	2328		
Db	1688	ACATTGAAGTGGACCCCTGACTGGGAGGCTGCCATCGCTATACAGATGCAACTTAAGAAATA	1747		
Qy	2329	ttttactcatgtttcccaagagtggtgcttgcgttgcgttgcgttgcgttgcgttgcgttgcgtt	2388		
Db	1748	TTTTGTCTCATGTTTCCCAAGAGTGTGTGCTTGTGATGAAGATCTCTTACTGTGTGCTTATA	1807		
Qy	2389	aagaaatgcacaaagctgtgatgaggtgaggtaccagtttcatatctagtagcaagacag	2448		
Db	1808	AAGAAATGTCACAAAGCTGTAATGAGGTGCAAGTACAAATTTTCATGTCCAGTACCAAGACAG	1867		
Qy	2449	tagtacaactcgtgtgacatagtttggaaaacaaagctccacagatctcagagatcttg	2508		
Db	1868	TAGTGAATTTGTGCGGTTCATAGTCTGGAAACCAAAATCCTACAAAGTGTCTGAGGACCTTG	1927		







Db 4088 CCTGGAGCACGTGTGCGTTCCACCATCCAGGCAATCGAAACCACTGTTGGGAGATGAAGAA 4147  
Qy 4717 aacctctgttggagcacttcaaaataggcagcataatgtctgtaaaagcattaatgcagt 4776  
Db 4148 AACCTCTATTTGGAGCACTTCAAAATAGACAGCATAGCGGTCTGAAGGCGCTAATGCAGT 4207  
Qy 4777 ttgcagttgcaagagattacctgtcctcaggtccctgatatcagaaacatctgttctc 4836  
Db 4208 TTGCAGTTGACAGAGGGCTACCTGCGCCTCAGGTCCTGATACACAAACATCTGGCTCGGC 4267  
Qy 4837 ttcatacagttgttcttcaataaatacagaagatacaccatgcctctgtctctatag 4896  
Db 4268 TCCTGTCACTATTCTTCCTAACCTGCACATCAGAAATACACAGGCTTCTGTCTGTGG 4327  
Qy 4897 atctgtttcatgttttgggtggtcgtgtgttagcatctcccatctgttattgggatgacc 4956  
Db 4328 ATCTCTTCCATGTTCTGTGTGGCGCAGTCTTAGCGTTCCCATCTCTTATTGGGATGACA 4387  
Qy 4957 ctgttgatcgcagccttcttcagttagttcttctcattataaccacctttatctctccatt 5016  
Db 4388 CCGTGGATCTGCAGCGCTGCGCACTTAGTCTTCATATAACACACCTCTATCTCTCCATC 4447  
Qy 5017 tgcacacatggcacatgcttcagatactacttaccagtagaca-----cag 5064  
Db 4448 TGATCACATGGCCACATGCTTCAGATACTCCCTTACACAGATACAGATCTGTCTCCAG 4507  
Qy 5065 gcctacccttctgcaggttcaagaagacagtgaagagctcattccgcattctcttct 5124  
Db 4508 GGCCGCGCTTGTGAGGTTGAAGAGGATAGTGAGGAGGCTCGTGTGCATCTGCTTCT 4567  
Qy 5125 ttgcagaattctcacaatatacagaagtggtccattgggtgtagattcctggtggtatt 5184  
Db 4568 TTGTGGAAGTGCAGCAGCACAGACGGCTTCACTGGGTGGCGGTCTCCCGCTGGTACC 4627  
Qy 5185 tgtgggtctcaagaatgcaatcccttcttctcgtctgctcattgttttcc 5244  
Db 4628 TGTGGCTCTCCCTGAGGAACGGCATCACCCCTTACCTCGGTGTGCTGCATCTGCTTTCC 4687  
Qy 5245 actatttacttgggttaactccgctgaggaactgcataccaaattctgcagaagagagt 5304  
Db 4688 ACTATTACTTGGAGTAGCTCCGCTGAAGAACTGTTGGCCAACTCTGCTGAAGAGAAAT 4747  
Qy 5305 acagtgcactctgtagctatctatttaccatacaaaattgttctcgtctcttccaggaat 5364  
Db 4748 TCAGTGCACCTGTAGCTATCTATCTTTACCCACAAATTTGCTGCTTTTCCAGGAAT 4807  
Qy 5365 attggatactgaagcccttgcagaggtggtgtagatcctccttactaaact 5424  
Db 4808 ATTGGGATACCATAAGGCCCTTACTACAGAGGTGGTGTGGAGATCCTGCCCTTACTCAAGT 4867  
Qy 5425 gtttgaagcaaaaaaacacccgtggtcaggtacccttagaaaaaataagtttgatagagc 5484  
Db 4868 CTTTGAAGCAGAAAGTCTGTGTCAGCTACCTACCTAGAAAAGAAATAGTTGATAGAGC 4927  
Qy 5485 ttctgtgactatagctgctccttgaaatcaagcttctcatttccaggtgcccacggtctg 5544  
Db 4928 TTCTCGAGGACTACAGCTGTCTCTTAATCAGGCTTCTCACTTTAGGTGTCCACGGTCTG 4987  
Qy 5545 cagatgtagcgaagacatcgtctcctgctcttctgtggtggtctatactatgtctc 5604  
Db 4988 CAGATGATGAGCAAAAGCATCTGTCTCTGTCTTCTGTGGGCCATCCTGTGTCTC 5047  
Qy 5605 agaacattgtgcccagaaattggaacggggaagaggttggaactgtgcatttttccag 5664  
Db 5048 AGAACATCTGTGGCCAAAGAAATAGTGAATGGGAAAGAGGTTGGAGCGTGGCTTTTCATG 5107  
Qy 5665 cacttctggtgagcccgagctgctcatttctcctaaataacagaaatgcccagtggtctc 5724  
Db 5108 GCGTCACTGTGGTCTGGAGCTGTGCATTTCTTCTAAATCCGAGAAATCCGAGGGTGGTCC 5167  
Qy 5725 tgggtgaaggttaagccagaggtgtgctatccagctccttacttgatgaatgag 5784  
Db 5168 TGGTGAAGAAAAAGCCAGAGCTGTGCTTACCCAGCCCTTACTTGGATGAATATGGAG 5227

Qy 5785 aaacgagaccctggcctgaagaggggcaaccccttcaatttatctctgagcggtatcgga 5844  
Db 5228 AAACAGACCCAGGGCTAAAGAGAGAAACCCACTTCATTATTCTTCGGGAGCGGTATCGGA 5287  
Qy 5845 agctccatttggctcggcaacaacactgcattatagaagattgctagagccaaagaga 5904  
Db 5288 AGCTGCATTTGGTCTGGCAACAGCACATGCTATTATAGAAGATTGTCGAGGCCAGAGA 5347  
Qy 5905 ctaatcagatgttatttggattcaactggcagttactgtgagctccaactctgcctcaag 5964  
Db 5348 CTAATCAGATGCTATTGGAATTAACGGCAGTTACTCTGAGCTTCAGTTCTGSCCTCAAG 5407  
Qy 5965 acaatcacaatgacagcagtagttaaagctgattcacaataattatgaaaaacttctcgagg 6024  
Db 5408 ACAATCATGAGTGAGCATCAATAATAAAGACTGATCTAAAAATCTAGAGAACTTCTGAGG 5467  
Qy 6025 gctggaaaagattgaggggtcttttgcctcgaatgcccaggttcaacttacaataaaat 6084  
Db 5468 ACGGGGGAAGTATTGGAGGGTCTTTTGATCTCATGTCCAGATTCACACATTAATAAAAT 5527  
Qy 6085 atttctaatggagctattgctttcaattagcaaacatgcttcacaggaataaa-aggac 6143  
Db 5528 ATTCTTAAATGAATATTGCTTTCAATTATCAACATAAGCTTCAGGGGAAAAACAAGAC 5587  
Qy 6144 atagatcaatctgt-----ttt 6160  
Db 5588 ATGATATTAAATGTTTATGTTCTAGAACACATAAAGAAATGCTTGTTCATCCAAGTGCTAT 5647  
Qy 6161 atgtcctagatttccaggaatttattcccttcaataattgtctcatttctatttattt 6220  
Db 5648 TTCTCTAATAATTCAGAAAACTCCTTTCCCTTCATAACTGTCCCTAGTTTCATTATCATAT 5707  
Qy 6221 catcacttgtagatgaagtcacgcacaaacagttgtagacatttattgtgttggttaac 6280  
Db 5708 CACCACCTGGTTAATGAGGTCACATTAGCATTTGTGGACATTTCTCCATCTGGCTAAC 5767  
Qy 6281 tcttctgcaatttgtatttgggtttt 6307  
Db 5768 ATCTCTGCACCTTGTATTGTTGGTGT 5794  
RESULT 4  
AK027803  
LOCUS Homo sapiens cDNA FLJ14897 fis, clone PLACE1004743, weakly similar  
DEFINITION to PROBABLE N-END-RECOGNIZING PROTEIN.  
ACCESSION AK027803  
VERSION AK027803.1 GI:14042751  
KEYWORDS oligo capping; fis (full insert sequence).  
SOURCE Homo sapiens placenta cDNA to mRNA, clone\_lib:PLACE1  
clone:PLACE1004743.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (sites)  
AUTHORS Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,  
Nishikawa,T., Nagai,K., Sugano,S., Aotsuka,S., Yoshikawa,Y.,  
Matsunawa,H., Ishii,S., Kawai,Y., Saito,K., Yamamoto,J.,  
Wakamatsu,A., Nakamura,Y., Nagahari,K., Masuho,Y. and Sasaki,N.  
NEDO human cDNA sequencing project  
TITLE Unpublished  
JOURNAL 2 (bases 1 to 2550)  
REFERENCE Isogai,T. and Otsuki,T.  
AUTHORS Direct Submission  
TITLE Submitted (10-MAY-2001) Takao Isogai, Helix Research Institute,  
Genomics Laboratory, 1532-3 Yana, Kisarazu, Chiba 252-0812, Japan  
JOURNAL (E-mail:genomics@hri.co.jp, Tel:81-438-52-3951, Fax:81-438-52-3952)  
COMMENT NEDO human cDNA sequencing project supported by Ministry of  
Economy, Trade and Industry of Japan; cDNA full insert sequencing;  
Research Association for Biotechnology; cDNA library construction;  
5'- & 3'-end one pass sequencing and clone selection; Helix  
Research Institute (supported by Japan Key Technology Center etc.)

FEATURES	source	BASE COUNT	ORIGIN
and Department of Virology, Institute of Medical Science, University of Tokyo.		830 a	708 t
Location/Qualifiers		958 g	454 c
1. .2550			
/organism="Homo sapiens"			
/db_xref="taxon:9606"			
/clone="PLACE1004743"			
/tissue_type="placenta"			
/clone_lib="PLACE1"			
/note="cloning vector: pME18SFL3"			
117. .>2550			
/note="unnamed protein product"			
/codon_start=1			
/protein_id="BAB55380.1"			
/db_xref="GI:14042752"			
/translation="MTIWEKEKLPPPELQIREKNERYCYVLFNDEHHSYDHVYLSIOR ALDCEAEQLQHTATDKGRRVAKAGAYACQAEAKEDKSHSENYSQHPLHVVEYLSQR EIMAHQKFAFLRLGSWNKIMSYSDRFQICQACLRPEEDSENPCILSRMLMDKALY KGARKILHELIFESFMEWEYKKLFAMEFYVYKIQKEYISDDHRSITSITALSYQM FVTPTLRLHIEQNTSVITEITLLEVLPEYLDNRNKNFYQGSQKLRGVAVICDL KYLISKPTIWTRELRAQFEGFRSLKILTCMQGMEETRRQVGHEIVDPDWEAIIA IOMOLNILLFMEWCAQDEBELLIVAYKCHKCAVMRCSTSFISSSYTVQSCGHSLET KSYRVEDLSVSLHPSRLAGLHVLSRLGAVSRLHEFVSDFEQVEVLVEYPLRCL VLVAQVVAEMRRNGLSLSQVFFYQDVKCREEMYDKDILMIQIGASLMDPNKFLFLY LQRYELAEAFNKTISTKQDLIQYNTLIEEMQLQVLIYIVGRIYVPGGVNVTKEEYTM RETILHLCITPEMPSIAKNLPENNETTGLENVINKVATFKPGYSGHGVNLLKNC LKDQNFVYHYSTQHSKAHEMKOKRRKOENKDEALPPPPPEFCFAFSKVINLLQD IMMYILRTVPERAIDTNSLWTCGMLQMAFHILALGLEEKQQLKAPBEETVDFYH KASRLGSSAMNIOMLLEKLGIPLQEGQDKDMITWILQMFDTVARLREKSLCIIVATTSS SESIKNDIEITHK"			

Query Match	40.4%	Score 2546.8	DB 9	Length 2550
Best Local Similarity	99.9%	Pred. No. 0		
Matches 2548	Conservative 0	Mismatches 2	Indels 0	Gaps 0
QY 1189	atcatgaacctggaagcgccagggtactataaagaagaattcacgcgtgcctgttgatgaag	1248		
Db 1	ATCATGAACCTGGGAAGCAGCGTACTATAAAGAGAATTCACGCTGTCCTGTAATGAAG	60		
QY 1249	aggtaattgtcccaagccaggaaaaatttcttcctcaatgataaaaatatgtcgtgagaatga	1308		
Db 61	AGSTPAATGTGCCAAGCCAGGAAAAATATTCTTCAGTGTAAAAATATGCCGTAGAAATGA	120		
QY 1309	ctatatgggaaggaagaaagaaactgcctcctgaactccagataaaggagaaaaatgaaa	1368		
Db 121	CTATATGGGAAGAGGAAAAAGAACTGCCTCTGAAC7CCAGATGAAGGAGAAAAATGAAA	180		
QY 1369	gatactattgtctcttttcaatgatgaacacattcatatgaccacgtcatatacacgcc	1428		
Db 181	GATACTATTGTGTCCTTTTCAATGATGAACACCACTTCATATGACCACGTATATACAGCC	240		
QY 1429	tacaaagagctcttgactgtgagctgcagagggccagttgcataccactgcccattgaca	1488		
Db 241	TACAAAAGAGCTCTTGACTGTGAGCTCCGAGAGGCCAGTTGCATACCAC7GCCATTGACA	300		
QY 1489	aagagggtcgtcgggctgttaaagcgggagcttatgctgcttgcaggagaagcaaggaag	1548		
Db 301	AAGAGGTCGTCGGGCTGTAAAGCGGGAGCTTATGCTGCTGCCAGGAAGCAAGGAAG	360		
QY 1549	atataaagagtcattcagaaaaatgtctcacaatccacttcattgtagaagtattacact	1608		
Db 361	ATATAAAGAGTCATTACAGAAAAATGCTCTCAACATCCACTTCATGTAGAAGTATTACACT	420		
QY 1609	cagagattatggctcattcagaaaaatttgcttcttgcttgcttgcttgatgaacaaaaatta	1668		
Db 421	CAGAGATTATGGCTCATCAGAAAAATTGCTTTGCGCTCTTGGTCTCTCGGATGAACAAAAATTA	480		
QY 1669	tgagctattcaagtgaacttaggcagatcttttgccaagcatgccttagagaagaacctg	1728		
Db 481	TGAGCTATTCAAGTGACTTTAGGCAGATCTTTTGCCAAAGCATGCTTTAGAGAACAACCTG	540		

Qy	1729	actcggaagaatccctgtctctctaagcaggttaatgctttgggagtgaagactttataaag	1789
Db	541	ACTCGGAAGAAATCCCTGTCTCTCATAAAGCAGGTTAATGCTTTGGGATGCAAGGCTTTATAAAG	600
Qy	1789	gtgcccgttaagatcccttcgatgaattgatcttcagcaggttttttttatggagatgggaataaca	1848
Db	601	GTGCCC GTTAAGATCCCTTCATGAANTGATCTTCAGCAGT TTTT TTTATGGAGATGGAAATACA	660
Qy	1849	aaaaactcttctgacttggaatttgtgaagtattataaaaactgacagaagaatatatca	1908
Db	661	AAAAACTCTTTGCTATGGAA TTTGTGAAGTATTTATAACAAC TGCAGAAAGNATATATCA	720
Qy	1909	gtgatgatcatgacagaagtatctctataactgcactttcagttcagatgtttactgttc	1968
Db	721	GTGATGATCATGACAGAAGTATCTCTATAACTGCAC TTTTCAGTTCAGATGTTACTGTTC	780
Qy	1969	ctactctggctgcacatcttattgaagacagaagtgttatctctgtctacttgaaactc	2028
Db	781	CTACTCTGGCTGCACATCTTATTGAAGACGAGAATGTATTCTCTGTCTATTACTTACTGAAACTC	840
Qy	2029	tgctagaagttttactctgacttgacagagaacaataaaattcaacttccaggtttata	2088
Db	841	TGCTAGAAGTTT TACTTGAGTACTTGCAGCAGGAACATAAATTCACCTCCAGGGTTATA	900
Qy	2089	gccaggacaattgggaagtgatatgcagtaatatgtgacctaaagtatatctctgatca	2148
Db	901	GCCAGGACAAATGGGAACAGTATATGCAGTAATATGTGACCTTAAAGTATATCCTGTATCA	960
Qy	2149	gcaaccaccaatattggacagaaaagattaaagaatgcagttcccttgaaggtttttogatctt	2208
Db	961	GCAAAACCCACAATATGGACAGAAAAGATTAAAGATGCGAGTTCCTTGAAGGTTTTCGATCTT	1020
Qy	2209	ttttgaagattcttacctgtatgcagggaatggaagaaatccgaagacaggttgggcaac	2268
Db	1021	TTTTGAAGATTCTTACTGTATGCAGGGAATGGAAGAAATCCGAAGACAGGTTGGGCAC	1080
Qy	2269	acattgaagtggatccctgatgtggaggtgccatttgctatcacagatgcaattgaaagata	2328
Db	1081	ACATTGAAGTGCATCCTGATTGGGAGGCTGCCATTGCTATACAGATGCAATTGAAGAATA	1140
Qy	2329	ttttactcatgttccaagagtggtgtgcttgtgatgaagaactcttacttgtggcttata	2388
Db	1141	TTTTTACTCATGTTCCAAGAGTGGTGTGCTGTGATGAAGAAC TCTTACTTGTGGCTTATA	1200
Qy	2389	aagaatgtcaaaaagctgtgatgaggtgcagtaaccagtttcatatctagttagcaagacag	2448
Db	1201	AAGAATGTCAAAAAGCTGTGATGAGGTGCAGTACCAAGTTTCATATCTAGTAGCAAGACAG	1260
Qy	2449	tagtacaatcgtgtgacatagtttggaaacaaagtcctacagagtatctcgaggtattctg	2508
Db	1261	TAGTACAATCGTGTGGACATAGTTTGGAAACAAGTCCCTACACAGTATCTTGAGGATCTTG	1320
Qy	2509	taagcatatcatctgcactctctaggaacctgtgcttggtcttcattgactgttttaagcaggc	2568
Db	1321	TAAGCATACATCTGCCACTCTCTAGGACCCCTTCTGCTGGTCTTCTATGTACGTTTAAAGCAGGC	1380
Qy	2569	tgggtgctgtttcaagactgcatgaatttgtgtcttttgaggactttcaagttagaggtac	2628
Db	1381	TGGGTGCTGT TTTCAAGACTGCGATGAATTTGTGTCTTTTTCAGGACTTTTCAAGTAGAGGTAC	1440
Qy	2629	tagtggaaatatcctttaagttgtctcgggtgttggtgtgccaggttgttgcgagatgtggc	2688
Db	1441	TAGTGGAAATATCCTTTACGTTGTCGTGGTGTGGTGTGCCCAGGTTGTTGCTGCAGATGTGGC	1500
Qy	2689	gaagaaatggactgtctcttattagccaggtgtttttattaccaagatgtttaagtgcagag	2748
Db	1501	GAAGAAATGGACTGCTCTCTATTAGCCAGGTGTTTATTACCAAGATGTTTAAAGTGCAGAG	1560
Qy	2749	aagaaatgtatgataaagata tcaatgtcttcaagatattggtgcacttttaattgagatcccca	2808
Db	1561	AAGAAATGTATGATAAAGATATCATCATGCTTCAGATTTGTTGCACTCTTTAATGGATCCCA	1620

Qy 2809 ataagttcttgaactgacttcagaggtatgaacttgcagaggttttaacaagacca 2868  
Db 1621 ATAAAGTCTTGTACTGTGACTTCAGAGGTATGAAGTTCGCCAGGCTTTTAAACAGACCA 1680  
Qy 2869 tatctacaaaagaccagatttgattaaacaataaatacactaataagaagaatgcttc 2928  
Db 1681 TATCTACAAAGACCAGGATTGTATTAACAATATATACACTATATAGAGAAATGCTTC 1740  
Qy 2929 aggtctcatctatatgttggtgagcgttatgtacctggagtggaagtgtgaccaag 2988  
Db 1741 AGGTCTCATCTATATGTGGGTGAGCGTTATGTACCTGGAGTGGAAATGTGACCAAG 1800  
Qy 2989 aagaggtcacaaatgagagaaatcattcactgcttttgattgaaccccatgccacagtg 3048  
Db 1801 AAGAGTCCACATGAGAGAAATCATTTACTTGTCTTTGGATTGAACCCATGCCACACAGTG 1860  
Qy 3049 ccattgcaaaaatttactgagaatgaaataatgaactggcttagagaatgtcataa 3108  
Db 1861 CCATTGCCAAAATTTACTTGGAATGAATATGAAGTGGCTTAGAGAAATGTCTATA 1920  
Qy 3109 acaagtgccacatttaagaacacaggtgtatcgagggccatggagtttatgaactaaaag 3168  
Db 1921 ACAAGTGGCCACATTTAAGAAACAGGTGTATCAGGCCATGGAGTTTATGAATAAAG 1980  
Qy 3169 atgaactcaatgaagaacttcaatgtacttttacttttacttaccataaaacccagcataga 3228  
Db 1981 ATGAATCTACTGAAGACTTCAATATGTACTTTATTTATCTACCCAAACCCAGCATAGCA 2040  
Qy 3229 aggtcgaaatgacagagaaagagaaacaaagaaagaaagaaagaaagaaagaaagaaag 3288  
Db 2041 AGGCTGAACATATGACAGAAAGAGAGAAACAGAAACAGAAACAGAAATGAAGCATTTGCCGC 2100  
Qy 3289 caccaccacctctgaattctgcctgcttttcagcaaaagtattaaccttctcaactgtg 3348  
Db 2101 CACCACCACCTCTGAATTTCTGCCCTGCTTTTCAGCAAAAGTGATTAACTTTCTCAACTGTG 2160  
Qy 3349 atatatgatgtacattctcaggaacgttatgttagcgggcaatagacacagattctaaact 3408  
Db 2161 ATATCATGTATGATCTCTCAGGACCGTATTTGAGCGGCAATAGACACAGATTCTAACT 2220  
Qy 3409 tgtgacccaagggatgtccaaatggcttttcataattctgcattgggtttactagaag 3468  
Db 2221 TGTGACCGAAGGATGCTCCAAATGGCTTTTCATATTTCTGCGATTTGGCTTTACTAGAAG 2280  
Qy 3469 agaagcaacagcttcaaaaagctctgagaagaagtaacatttgactttttactataagg 3528  
Db 2281 AGAAGCAACAGCTTCAAAAAGCTCTGAAGAGAGTAACATTTGACTTTTATCATAGG 2340  
Qy 3529 cttaagattgggaagttcagccatgaatatatacaaatgcttttgaaaaactcaaggaa 3588  
Db 2341 CPTCAAGATTGGGAAGTTTCAGGCATGAATATACAAATGCTTTTGGAAAACCTCAAGGAA 2400  
Qy 3589 ttccccagttagaagccagagacatgataacgtgatacttcagatgtttgacacag 3648  
Db 2401 TTCCCCAGTTTGAAGCCAGAGGACATGATAACGTGGATCTTCAGATGTTTGCACACAG 2460  
Qy 3649 tgaagcgttagaagaaaaattgttttaattgtagcaaccacatcaggatcggaatcta 3708  
Db 2461 TGAAGCGGTAGAGAAAAATCTTGTTPAATTTGTAGCAACACATCAGGATCGGAATCTA 2520  
Qy 3709 ttaagaatgatgagattactcatgataaag 3738  
Db 2521 TTAAGAATGATGAGATTACTCATGATAAG 2550  
RESULT 5  
AR030785 1001 bp DNA PAT 29-SEP-1999  
LOCUS  
DEFINITION Sequence 2 from patent US 5861312.  
ACCESSION AR030785  
VERSION AR030785.1 GI:5943999  
KEYWORDS  
SOURCE Unknown.

ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 1001)  
AUTHORS Varshavsky,A. and Kwon,Y.Tae.  
TITLE Nucleic acid encoding mammalian UBRL  
JOURNAL Patent: US 5861312-A 2 19-JAN-1999;  
FEATURES Location/Qualifiers  
source  
1..1001  
/organism="unknown"  
BASE COUNT 363 a 186 c 205 g 247 t  
ORIGIN  
Query Match 15.8%; Score 996.2; DB 6; Length 1001;  
Best Local Similarity 99.7%; Pred. No. 2.6e-240;  
Matches 998; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
Qy 2799 atggatcccaataaagttctgttacttggttacttcaaggtatgaacttgcgaggtttt 2858  
Db 1 ATGGATCCCAACAAGTTCTTTGTACTTGGTACTTCAGAGGTATGAAGTTCGCGAGGCTTTT 60  
Qy 2859 acaagaccatatctcaaaaagaccaggtattgtattaaacaataatacacactaataaga 2918  
Db 61 ARCAAGACCATACTACAAAAGACCAGGATTTGATTAACAATATATACACTAATAGAA 120  
Qy 2919 gaaatgctcaggtcctcctatctatgttgggtgagcgttatgtacctggagtggaagt 2978  
Db 121 GAAATGCTTCAGTCCCTCATCTATTTGTGGTGAGCGTTATGTACCTGGAGTGGGAAT 180  
Qy 2979 gtgaccaaaagagaggtcacaaatgagagaataatcattcacttctgttgcattgaacccatg 3038  
Db 181 GTGACCAAGAGAGAGTCAATGAGAGAAATCATTCACTTGCTTTGCATTTGAACCCATG 240  
Qy 3039 ccacacagtgccattgccaaaaatttacctgagaatgaatgaatgaatgaatgaatgaatga 3098  
Db 241 CCACACAGTGCCTATTTGCCAAAAATTTACCTGAGAAATGAATGAATGAATGAATGAATGA 300  
Qy 3099 aatgtcataaaaaagtgccacatttaagaacccaggtgtatcagggccatggagtttat 3158  
Db 301 AATGTCTATAAACAAGTGGCCACATTTAAGAAACCCAGGTGTATCAGGCCATGGAGTTTAT 360  
Qy 3159 gaactaaagatgaatcactgaaagacttcaatgtacttcttacttacttacttacttacttact 3218  
Db 361 GAACTAAAAGATGAATCACTGAAAGACTTCAATATGTACTTTTATCTATTACTCCANAACC 420  
Qy 3219 cagcatagaagctgaacatatgcaagaagaagaaggaagaacaagaacaagaagatgaa 3278  
Db 421 CAGCATAGCAGCTGAACATATGCAGAAAGGAGAAACAAAGAAACAAAGAAAGATGAA 480  
Qy 3279 gcaatgccccacacacacccctcctgaattctgcctcgtcttccagcaaaagtgttaacctt 3338  
Db 481 GCATTGCGGCCACCACCCCTCTGAAATTTCTGCTCTGCTTTTCAGCAAGATGATTAACTT 540  
Qy 3339 ctcaactgtgatcatcatgatgtacattctcaggaacgtatttgagcggggaatagacaca 3398  
Db 541 CTCAACTGTGATATCATGTATGTACTCTCAGGACCGCTATTGAGCGGGCAATTAAGACA 600  
Qy 3399 gatctaaacttgaccggaaggtctcctcaaatggcttttccatattcttgcgcatgggt 3458  
Db 601 GATTCTAACTTTGGACCCGAAGGATGCTCCAATGGCTTTTCATATTTCTGGCAATTTGGGT 660  
Qy 3459 ttactagaagagcaaacagcttcaaaaagctcctgaagaagaagtaacatttgactttt 3518  
Db 661 TTACTAGAGAGAAAGCAACAGCTTCAAAAAGCTCTCAAGNAGAAAGTAACATTTGACTTT 720  
Qy 3519 tatcataaggttcaagattgggaagttcagccatgaatatacaaaatgcttttggaaaaa 3578  
Db 721 TATCATAAGGCTTCAAGATTGGGAAGTTCAGCCATGAATATACAAATGCTTTTGGAAAAA 780  
Qy 3579 ctcaaaaggaattccccagtttagaaggccagagacatgataacgttgatacttcagatg 3638  
Db 781 CTCAAAAGGAATCCCCAGTTAGAAAGGCCAAGGACATGATTAACGTGGATACTTCTCAGATG 840

Qy	3639	tttgacacagtgaagcgtattaaagagaaaaatcttgttttaattgtagcaaacacatcaggga	3698
Db	841	TTTGACACAGTGAAGCGGATTAAAGAGAAAAATCTGTGTTAAATTGTAGCAACACATCAGGA	900
Qy	3699	tcggaactttaagaatgatgagattactcatgataaagaaaaagcagaacgaaaaaga	3758
Db	901	TCGGAATCTATTAAAGATGATGAGATTACTCATGTAAGAAAAAGCAGACGAAAGA	960
Qy	3759	aaagctgaagctgtaggctacatcgccagaagatcatggc	3799
Db	961	AAAGCTGAAGCTGTAGGCTTCATCGCAGAAAGATCATGGC	1001
RESULT 6			
AR121464	AR121464	1001 bp	DNA
LOCUS	Sequence 2 from patent US 6159732.		PAT
DEFINITION	AR121464		
ACCESSION	AR121464		
VERSION	AR121464.1	GI:14105040	
KEYWORDS	Unknown.		
SOURCE	Unknown.		
ORGANISM	Unclassified.		
REFERENCE	1 (bases 1 to 1001)		
AUTHORS	Varshavsky, A. and Kwon, Y. Tae.		
TITLE	Nucleic acid encoding mammalian Ubr1		
JOURNAL	Patent: US 6159732-A 2 12-DEC-2000;		
FEATURES	Location/Qualifiers		
source	1..1001		
BASE COUNT	363 a 186 c 205 g 247 t		
ORIGIN			
Query Match 15.8%; Score 996.2; DB 6; Length 1001;			
Best Local Similarity 99.7%; Pred. No. 2: 6e-240;			
Matches 998; Conservative 0; Mismatches 3; Indels 0; Gaps 0;			
Qy	2799	atggatcccaataagttcttctgtactgtacttcagaggatgaacttgcgagagctttt	2858
Db	1	ATGGATCCCAACAAGTCTTGTACTGTGTTACTTCCAGAGGTATGAACCTGCCGAGGCTTTT	60
Qy	2859	acaaagaccatatctacaaagaccaggatttgattaaaaaatataacactaatagaa	2918
Db	61	AAACAAGACCATATCTACAAAGACACAGAGTTGATTAAACAATAATAACTAATAAGAA	120
Qy	2919	gaatgcttcagctccatctatattgtggtagcggtattgtacctgaagtgggaat	2978
Db	121	GAATGCTTCAGGTCCTCATCTATATTGTGGTGAGCGTTATGTACTCTGGAGTGGGAAT	180
Qy	2979	gtgaccaaagaagagggtccaaatgagagaaatcattccactgtcttgcattgaaccatg	3038
Db	181	GTGACCAAAAGAGAGGTCACAATGAGAGAAATCATTTCATTCGTTTGCATTGAACCATG	240
Qy	3039	ccacacagtgccattggccaaaaatttacctgagaatgaaataatgaaactgggttagag	3098
Db	241	CCACACAGTGCCATTGGCCAAAAATTTTACCTGAGAATGAAAATAATGAAATGGCTTAGAG	300
Qy	3099	aatgtcataaacaaggtggccacatttaagaaccaaaggtgtatcaggccatgagtttat	3158
Db	301	AATGTCTATAAACAAAGTGGCCACATTTTAAAGAACACAGGTGTATCAGGCCATGGAGTTTAT	360
Qy	3159	gaactaaaagatgaatcactgaaagacttcaatgtacttttacttattcattactcaaaacc	3218
Db	361	GAACTAAAGATGAATCATTGAAGACTTCAATATGTACTTTTATCATTTACTTCCAAACC	420
Qy	3219	cagcatagcaaggtcgaacatatgcagaagaaaaaggagaaaaaagaaaaaagaatgaa	3278
Db	421	CACATAGCAAGGCTGAACATATGCAGAGAAAAGGAGAAAACAGAAAACAAGATGAA	480
Qy	3279	gcattgccgccaccacacacctcctgaattcgcctgtcttcagcaaatgtataacctt	3338
Db	481	GCATTGCCGCCACACACACCTCCTGAATTCGCTCTTTCAGCAAGGTGATTAACTT	540

Qy	3339	ctcaactgtgatcatcatgtagatcattctcaggaccgtattttagcgcggaatagacaca	3398
Db	541	CTCAACTGTGTATATCATGATGATACATTTCTCAGGACCGGTATTTGAGCGGCAATAAACACACA	600
Qy	3399	gattctaacttgttgaccgaaggatgctccaaatggctttttcattatcttggcattgggt	3458
Db	601	GATTCTAACTTGTGGACCGAAGGGATGCTCCAAATGGCTTTTCATATTTCTGGCATTGGGT	660
Qy	3459	ttactagaagaagaacacagcttccaaaagctcctctgaagaagaagtaacatttgacctt	3518
Db	661	TTACTAGAAGAAGAAGCAACACAGCTTCAAAAAAGCTCTCGAAGAAGAAGTAACATTTGACTTT	720
Qy	3519	tatcataagcttcaagatgtgggaagttcagccatgaatatacaaatgtcttttggaaaaa	3578
Db	721	TATCATTAAGGCTTCAAGATTGGGAAGTTCAGCCATGAATATACAAATGCTTTTGGAAAAA	780
Qy	3579	ctcaaaggaattccccagtttagaaggccagaagacatgataacgttgatacttcagatg	3638
Db	781	CTCAAGGAATTTCCCAAGTTTAGAAGGCCAGAAAGACATGATAACGTGATACTTTCAGATG	840
Qy	3639	tttgacacagtgaagcattaaagagaaaaatcttgttttaattgttagcaaacacatcagga	3698
Db	841	TTTGACACAGCTGAAGCGATTAAAGAGAAAAATCTTGTTTAATTGTGAGCAACACATCAGGA	900
Qy	3699	tcggaatcttaagaatgatgagattcattcatgataaagaaaaagcagaacgaaaaaga	3758
Db	901	TCGGAATCTATTAGAATGATGAGATTACATGATTAAGAAAAAGCAAGCAAGCAAGAAAAA	960
Qy	3759	aaagctgaagctgtaggctacatcgccagaagatcatg9c	3799
Db	961	AAAGCTCAAGCTGTAGGCTTCATCGCCAGAAAGATCATGGC	1001
RESULT 7			
AF061556			
LOCUS			
DEFINITION			
AF061556 999 bp mRNA PRI 10-JUL-1998			
Homo sapiens ubiquitin-protein ligase E3-alpha (UBR1) mRNA, partial cds.			
AF061556			
AF061556.1 GI:3170888			
human.			
ORGANISM			
Homo sapiens			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
REFERENCE			
AUTHORS			
Kwon,Y.T., Relts,Y., Fried,V.A., Hershko,A., Yoon,J.K., Gonda,D.K., Swanon,P., Copeland,N.G., Jenkins,N.A. and Varshavsky,A.			
TITLE			
The mouse and human genes encoding the recognition component of the N-end rule pathway			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
Submitted (24-APR-1998) Division of Biology, 147-75, California Institute of Technology, 1200 E. California Bl., Pasadena, CA 91125, USA			
FEATURES			
source			
1. . .999			
/organism="Homo sapiens"			
/db_xref="taxon:9606"			
/map="15q15-q21.1"			
/chromosome="15"			
1. . .>999			
/gene="UBR1"			
1. . .>999			
/gene="UBR1"			
/note="UBR1p; recognition component of the N-end rule pathway"			
/product="ubiquitin-protein ligase E3-alpha"			
/codon_start=1			

```
/protein_id="AAC39845.1"
/db_xref="GI:3170889"
/translation="MDPNKFLLLVLQRYELAEAFNKTISTQDDLIKQYNTLIEMLQ
VLIVIGBRYVPGVGNVTKKEVTMRIEIIHLICIEPMPSAJAKNIPENNNETGLENV
INKVATFKKPGVSGHGVYELKDFNMFYHYSKTQHSKAEHMQKRRKQENKDE
ALPPPPPEFCFAPSKNVTLNCDIMMYILRTVFERAINTDSNLTWTEGKLOMAFHILA
LGLLEKQOLQKAPEEVTFDFYHKASRLGSSAMNIOMLLEKLGIGIPOLEGOKDITW
ILQMPDVKRLREKSLIVATISSESINKDEITHDKEAERKRAEAARLHRQIM"
BASE COUNT 363 a 185 c 204 g 247 t
ORIGIN

Query Match 15.8%; Score 994.2; DB 9; Length 999;
Best Local Similarity 99.7%; Pred. No. 8.2e-240;
Matches 996; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2799 atggatcccaataagtcttgtaactgtacttcagagatgaacttcagagcgtttt 2858
|||||
Db 1 ATGGATCCCAACAGTCTTGTTACTGGTACTTCAGAGGTATGAACCTGCCGAGGCTTTT 60
|||||

Qy 2859 aacaagaccatatctacaaaagaccaggattgtgataaacaataataacactaatagaa 2918
|||||
Db 61 AACAGACCATATCTACAAAACACAGGATTTGATTAACATATATACATATAGAA 120
|||||

Qy 2919 gaaatgcttcaggtcctcatctattgtgggtgagcgttatgtacctggagtggaat 2978
|||||
Db 121 GAAATGCTTCAGTGCTCATCTATATATTGTGGTGAGCGTTATGTACCTGGAGTGGGAAT 180
|||||

Qy 2979 gtgacaaaagaaggtcaaatgagagaaatcattcacttgccttgcattgaacccatg 3038
|||||
Db 181 GTGACCAAGAAGAGGTCAATATGAGAGAAATCATTCACCTGCTTGGCATTAACCCATG 240
|||||

Qy 3039 ccacacagtgccatggccaaaatttacctgagaaatgaataataatgaactggttagag 3098
|||||
Db 241 CCACACAGTGCCATGTCGCAAAAATTTACCTGAGATGAATAATGAACACTGGCTTAGAG 300
|||||

Qy 3099 aatgtcataacaaagtgccacatttaagaacacagggtgtatcagggccatggatttat 3158
|||||
Db 301 AATGTATATAACAAAGTGCCACATTTTAAGAAACACAGGTGTATCAGGCCATGAGTTAT 360
|||||

Qy 3159 gaactaaaagatgaatcactgaaagacttcaatatgtacttttatactactccaaacc 3218
|||||
Db 361 GAACTAAAGAGTGAATCACTGAAAGACTTCAATATGTACTTTATCATTTACTCCAAACC 420
|||||

Qy 3219 cagcatagcaagcgtgaacatatgcagaagaaagagagaaacaaagaaacaaagatgaa 3278
|||||
Db 421 CAGCATAGCAAGCTGAACATATGCAGAAAGAGAGAAACAAAGAAACAAAGATGAA 480
|||||

Qy 3279 gcattgcgcgccaccacacctcctgaattctgccttgccttcagcaagtgattacacctt 3338
|||||
Db 481 GCATTGCCGCCACCACCCTCCTGAATTCGCCCTGCTTTTCAGCAAAAGTATTAACCTT 540
|||||

Qy 3339 ctcaactgtgatcatgatgatctctcaggaacogtatttgagcgggcaatagacaca 3398
|||||
Db 541 CTCAACTGTGATATCATGATGTACATTTCTCAGACCGTATTTGAGCGGCAATAAACACA 600
|||||

Qy 3399 gattctaaactgtggaccgaagggatgctccaaaaggcttttcataattcttgccatgggt 3458
|||||
Db 601 GATTCTAACTTGTGACCGCAAGGGATGCTCCAAATGGCTTTTCATATTCGCGATTTGGGT 660
|||||

Qy 3459 ttactagaagaagaacacagcttcaaaaagctccttgaagaagaagtaacatttgacttt 3518
|||||
Db 661 TTACTAGAGACAGACACAGCTTCAAAAAGCTCCTGAGAGAGAAGTACATTTGACTTT 720
|||||

Qy 3519 tatcataagcgttcgaagtgggaagttcagccatgaatatcaaaatgcttttggaaaaa 3578
|||||
Db 721 TATCATAAGGCTTCAAGATTGGAAGTTCCAGCCATGAATATACAAATGCTTTTGGAAAAA 780
|||||

Qy 3579 ctcaagaagaattccccagtagaagccagaagggacatgataacgtggatacttcagatg 3638
|||||
Db 781 CTCAAAGGAATTTCCCAGTTAGAAGGCCAGAGGACATGATACCGTGGATCTTCAGATG 840
|||||

Qy 3639 tttagacacagtgaaagcagtaagaagaaaaatctgttttaattgtagcaacccacatcagga 3698
|||||
```

```
|||||
Db 841 TTTGACACAGTGAAGCGATTAAAGAAAAATCTGTTTAATTGTAGCAACCATCATCAGA 900
|||||

Qy 3699 tcggaatctattaagaatgatgatgattactactcatgataaaaaagcagaaacgaaaaa 3758
|||||
Db 901 TCGGAATCTATTAAAGATGATGAGATTACTCATGATAAAGAAAAAGCAGAACGAAAAAGA 960
|||||

Qy 3759 aaagctgaagctcctaggctacatcgccagaagaatcatg 3797
|||||
Db 961 AAAGCTGAAGCTGTAGCTTCATCGCCAGAAGATCATG 999
|||||

RESULT 8
AB002347
LOCUS
DEFINITION Human mRNA for KIAA0349 gene, partial cds. PRI 13-FEB-1999
ACCESSION AB002347
VERSION AB002347.1 GI:2224638
KEYWORDS KIAA0349.
SOURCE Homo sapiens male brain cDNA to mRNA, clone_lib:pBluescriptII SK
plus clone:HG1561.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 6158)
AUTHORS Nagase,T., Ishikawa,K., Seki,N., Nakajima,D., Ohira,M.,
Miyajima,N., Kotani,H., Nomura,N. and Ohara,O.
Miyajima,N., Kotani,H., Nomura,N. and Ohara,O.
TITLE Direct Submission
JOURNAL Submitted (28-MAR-1997) to the DDBJ/EMBL/GenBank databases. Nobuo
Nomura, Kazusa DNA Research Institute, Gene Structure 1; 1532-3
Yana, Kisarazu, Chiba 292, Japan (E-mail:cdmainfo@kazusa.or.jp,
URL:http://www.kazusa.or.jp, Tel:+81-438-52-3930,
Fax:+81-438-52-3931)
REFERENCE 2 (sites)
AUTHORS Nagase,T., Ishikawa,K., Nakajima,D., Ohira,M., Seki,N.,
Miyajima,N., Tanaka,A., Kotani,H., Nomura,N. and Ohara,O.
TITLE Prediction of the coding sequences of unidentified human genes.
VII. The complete sequences of 100 new cDNA clones from brain which
can code for large proteins in vitro
JOURNAL DNA Res 4 (2), 141-150 (1997)
MEDLINE 9734984
FEATURES
source
Location/Qualifiers
1..6158
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HG1561"
/clone_lib="pBluescriptII SK plus"
/sex="male"
/tissue_type="brain"
1..3828
/gene="KIAA0349"
<...3828
/gene="KIAA0349"
/codon_start=1
/protein_id="BAA20806.1"
/db_xref="GI:2224639"
/translation="LKYVLISKPTENSDELROKLEGFDAFLLEKLMQCGMDPIITROV
GHIEMEPEWEARATLQMLKTHVISMQDCASDEKVIIEAYIKKCLAVLMQCHGGYTD
GQPIITLSTCGHVSFETIRYCSQEKVSIHLPLVSKLLAGLHLLSKSEVAYKFPPELLPL
SELPPLMLIEHPLCLVLCAQVHAGMWRNGFSLVNOIYYHHVVKRRFMDPDVNL
QTGVSMDPNHFLMILSRFELYQIFSTPDYGRKFSSEITHKDVQNNNTLIDEMLYL
ITLVGERFSPGVGVNATDEIKREIILHOLSKPMASSELVLSLPDENKETGMSVLI
EAVAHFKPGLTGRMYELKPECAKEFNLYFFHFSRAEQSKAEQAKLRQNRDFA
LPPVLPFCPLFASLVNIIQSDVLMICMTGILQWAVEHNGYAWSEMLQRLVLIGH
ALQOEKHLENVTTEHVFTTQKISKGEAPKNSPSILAMLETQNAIYLVHVKDM
IRWILKTFNAVKMRESSTPTSPAETEGTITMEESRDKDKAEKRAEIAARLREKIM
AOMSEORHFIIDENKELFOOTLELDSAVLDHSPVADMTLTALGPTQTVPEQRQ
FYTCILQOEQEVKVFESRAWLVAAAFVQSRVLSKRSKEIQDPEKYDPLPMHPDLSG
THYSSCGHIMHACHQRYFDSVOAKEQRRQORLUHTSYDVENGEFLCPLCECLSNV
IPLLPPRFINNRLFSQPNLTQWIRTIISQIKALQFURKEESTPNNASTKNSNV
DELQPEGRFPDPRKIPYSESIKEMLTITGTATYKVLKLVHNEEDPRVPMCGSC
ATTQISIERILSDEKPLGFLPCRLDCLSLRTFRAAAHWTVASVSVQGHCKLFA
SLVPNDSHEELPCILDIDMFHLVLGLVLAFFPALQCQDFSGISLTGDLHFHLVTMAH
```

IIQILLTSCTEENGMDENPPCEESAVILALYKTLHOYTGSLSALKEIPSGHWLRSVRA  
GIMPELKCSALFPHLYNGVSPDPIDQVPGTSHFEHLCSYLSLPNNLICFQENSETWN  
SLIESWRNSENKRLNGERDAIRYPRESKNLINUPEDYSSLINQASNFSCPKSGGDK  
SRAPTCLVCGSLICSQSYCCOFELEGACTAFYSCGSGVGFILVRRECQVLF  
AGTKGCFYSPYLDYGETDQGLRRGNPLHLCKERFKKIQKLWHQHSVTEIGHAE  
ANOTLVGIDWOHL"

BASE COUNT 1832 a 1199 c 1320 g 1807 t  
ORIGIN

Query Match 15.6%; Score 981; DB 9; Length 6158;  
Best Local Similarity 55.9%; Pred. No. 2.1e-236;  
Matches 2157; Conservative 0; Mismatches 1610; Indels 90; Gaps 12;

QY 2130 ctaaagatatctgcatcagcaaacccacaataatgagcagagaagattaaagatgcagttc 2189  
DB 1 CTCAAGTATGTGTTAAATAGCAAACTGTAATGGTGCAGATGAGTGGAGCAAGTTC 60  
QY 2190 cttgaaggttttcgatctctttttgaagattcttacctgtatgcaggggaatggagaataac 2249  
DB 61 CTAGAAGGGTGTGATGCCCTTTTGGAAATTAATAAATGTATGCAAGGAATGGATCCAATT 120  
QY 2250 cgaagacaggttgggcaacacattgaagtggaacctgattggagagctgccattgctata 2309  
DB 121 ACACGTCAAGTAGGACACATATTGAATGGAAACACAGAGTGGGAAGCAGCCCTCACACTA 180  
QY 2310 cagatgcaattgaagaatatcttactcatgttccaagagtggtgtcgttgatgaagaa 2369  
DB 181 CAAATGAATTAACACATGTCAATTTCAATGTATGCAGGACTGGTGTCTTCAGATGAAAA 240  
QY 2370 ctcttacttggcttataaagaatgcacaaagctgtgatgaggtgcagaccagtttc 2429  
DB 241 GTGTTAATCGAAGCTTTACAGAAATGCTCGCTGTACTGATGCAGTGTCAATGGTGGTTAT 300  
QY 2430 atatctagtagcaagacagt--agtcacaatcgttggacacatagtttggaaacaaagttcc 2486  
DB 301 ACTGATGGTGACAGCCATCACACTAAGCAATTTGGACATTCAGTGGAAATATACAGA 360  
QY 2487 tacagagtactgagatcttgaagcatacatctgccactctctaggaaccttgcctggt 2546  
DB 361 TACTGTGTTTCCCAAGAAAAAGTTAGCATTCACCTCCCACTTCCTCGCTTACTTGCAGGT 420  
QY 2547 ctctcatgacgtttaagcaggtcgtgctgttcttaagactgcataagatttgctctttt 2606  
DB 421 TTACATGTATTATTAAAGCAAAAGTGAAGTGGCATATAAATTTCCAGAGCTCCTACCTCTA 480  
QY 2607 gaggacttccaagtagagtagtactagtggaatctcttactgctgtggtgtggttggc 2666  
DB 481 AGTGAACCTTAGCCCACTTGTGATAGAACACCCCTCTTAGATGCTCTTGTCTGTGTGCC 540  
QY 2667 caggttgttgcgtgagatgtggcgaagaatggactgctcttattagcaggtgttttat 2726  
DB 541 CAAGTACATGCCGAATGTGGAGAAGAAATGGTTCCTCTAGTAACACAGATTTATTAC 600  
QY 2727 taccagaagtgttaagtgcagagaagaatgtatgataaagatatcatcattgcttcagatt 2786  
DB 601 TACCATAATGTGAAATGCAGACGTGAGATGTTTGACAAGGATGTAGTAATGCTTCAGACA 660  
QY 2787 ggtgcattttaatggatcccaataagttcttcttactggtacttcagaggtatgaactt 2846  
DB 661 GGTGCTCCATGATGCCAATCAATTTCCCTGATGATCATGTCTAGCCGCTTTTGAACCT 720  
QY 2847 gccaggtcttttaacaagaccatatctacaaa-----gaccag 2885  
DB 721 TATCAGATTTTTCAGTACTCCAGACTATGGAAGAAAGATTTAGTTCTGAGATTACCCATAAG 780  
QY 2886 gatttgattaaacaatatatacactaatagagaagaatgcttcaggtccctcatctatt 2945  
DB 781 GATGTGTTTCAGCAGAACAACTACTCTAATAGAGAATGCTATACCTCATTAATGCTT 840  
QY 2946 gtcgggtgagcgttatgtacctggagtgggaaatgtgacccaaagaaggttcacaaatgaga 3005  
DB 841 GTTGGAGAGAGATTTAGTCTCTGGAGTTGGACAGGTAAATGCTACAGATGAAATCAAGCGA 900

QY 3006 gaaatcattcactgcttcttgattgaaacccatgcccacacagtgccattgcccacaaattta 3065  
DB 901 GAGATTATCCATCATGTTGAGTATCAAGCCTATGGCTCATAGTGAATTGGTAAAGTCTTTA 960  
QY 3066 cctgagaatgaaaaataatgaactgcttagagaagtgcataaaacaagaatggccacattt 3125  
DB 961 CCTGAAGATGAGAAACAGGAGACTGGCATGGAGTGATGATCGACGATTTGCCCTTTTC 1020  
QY 3126 aagaaccaggtgtatcagggccatggagtttatgaactaaagaatgaatcactcagaaagac 3185  
DB 1021 AAGAAACCTGGATTAAACAGGACGAGGCATGTATGAATCGAAACCAAGATGTGCCAAGAG 1080  
QY 3186 ttcaatatgtacttttatcattccaaaacccagcagtagcaaggtcgaacatatgcag 3245  
DB 1081 TTCAACTGTGATTCTATCTACTCTTTCAAGGGCAGAACAGTCCCAAGGCAGAGAAGCGCAA 1140  
QY 3246 aagaaaaggagaaaaaacaagaataaagatgaagcattgcgcgccaccaccactcctgaa 3305  
DB 1141 CGGAAATTGAAAGACAAAAATAGAGAAGATACAGCACTCCCACTCCGGTGTGGCTCCA 1200  
QY 3306 ttctgcctcgtcttccagcaaatgatttaacctctcaactgtgatatcatgatgtacatt 3365  
DB 1201 TTCTGCCCTCTGTTTGCAGGCTGGTTAAACATTTTGCAGTTCAGATGCTCATGTTGTGCATC 1260  
QY 3366 ctccagggaccgtatttgagcgggcaatagacacagattcttaacttggaccgaaaggtg 3425  
DB 1261 ATGGGAACAATTCTGCAATGGCTGTGGAAACATAATGATATGCCCTGGTCAGAGTCCATG 1320  
QY 3426 ctccaaatggcttttctaatcttctggcattgggttttactagaagaagaacaacagcttaa 3485  
DB 1321 CTGCAAAAGGGTGTACATTTAAATGGCATGGCACTACAGAAAGAAAAACAACATTAGAG 1380  
QY 3486 aagctcctcaag--aagaagtaacatttgacttttatcataagccttcaagattggaa 3542  
DB 1381 AATGTCAGGAGAGACATGTAGTAACTTTACCTTCACTCAGAAATATCAAAACCTGGT 1440  
QY 3543 agttcagcccatgaat-----atacaaatgcttttggaaaaactcaaaaggaattccc 3593  
DB 1441 GAAGCGCCAAAAAATTCCTCTAGCATACTAGTATGCTGGAAACACTACAAAAATGCTCCC 1500  
QY 3594 cagttagaagggccagaaggacatgataaacgtggatactctcagatgttttgacacagtgaa 3653  
DB 1501 TACTAGAAGTCCCAAAAGACATGATTCGGTGGATATTGAAGACTTTTAAATGCTCTAAA 1560  
QY 3654 cgattagaagaataatcttctttaaattgtagcaaccacatcagatcggaatcttattaa 3713  
DB 1561 AAGATGAGGAGA-----GTTCACTTACCAGTCCCGTGGCAGACAGAGAAGAACCAT 1614  
QY 3714 aatgatgagattactcatataaagaaaaagcagaacgaaaaagaaagctgaaagctgct 3773  
DB 1615 ATGGAAGAGAGTTCAAGGGCAAAAGACAAAGCTGAGAGGAAGAAAAAGCAGAGATTGCC 1674  
QY 3774 aggtacatgcgcgaagatcatggtcagatgtctgcttacagaanaaacctcattgaa 3833  
DB 1675 AGACTGGCAGAGAAAAAGATCATGGCTCAGATGTCTGAATGTCAGCGCATTTTATTGAT 1734  
QY 3834 actcaataactcatgtatgacaatacatcagaataatgcctgggaagaagattccattatg 3893  
DB 1735 GAAACAAAGAACTCTTTTCAGCAGACATTAGAACTGGATGCTCAACCTCTGCTGTTCT 1794  
QY 3894 gaggaagagagcaccgccagtcagtgactactcttagaattgctttgggtcctaaacgg 3953  
DB 1795 GATCA-----TAGCCCTGTGGCTTCAGATATGACACTTACAGCACTGGGTCACACAA 1848  
QY 3954 ggtccatctgttactgaaaaggaggtgctgacgtgcatccttttggcaagaagaacagag 4013  
DB 1849 ACTCAGCTTCTCGAACAAGACAATTCGTTACATGATATTGTGTCAAGAGAGCAAGAA 1908  
QY 4014 gtgaaaaatagaaaaataatgcatcattggtattatcgccctgtgtccagaataatctactcctta 4073  
DB 1909 GTTAAAGTGGAAACAGCGGCAATGGTCTTGGCAGCAATTTGTTTCAGAGATCAACTGTATTA 1968





Genethon.  
Direct Submission  
Submitted (05-DEC-1994) Genethon, B.P. 60, 91002 Evry Cedex France.  
E-mail: Nucle@genethon.fr  
REFERENCE  
3 (bases 1 to 818)  
AUTHORS  
Chiannilkulchai, N., Pasturaud, P., Richard, I., Auffray, C. and Beckmann, J. S.  
TITLE  
A primary expression map of the chromosome 15q15 region containing the recessive form of limb-girdle muscular dystrophy (LGMD2A) gene  
JOURNAL  
Hum. Mol. Genet. 4 (4), 717-725 (1995)  
MEDLINE  
95359980  
FEATURES  
Location/Qualifiers  
1..818  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="15"  
/clone\_lib="CEPH YAC clone 774G4 was used for cDNA selection"  
/map="15q15"  
/tissue\_type="fetal muscles"  
BASE COUNT 225 a 177 c 195 g 221 t  
ORIGIN  
Source  
Query Match 12.1%; Score 764.2; DB 9; Length 818;  
Best Local Similarity 98.2%; Pred. No. 9.2e-182;  
Matches 783; Conservative 0; Mismatches 13; Indels 1; Gaps 1;  
QY 5302 agtacgtcacctgtactatctattacctaacaattgttctcgtctctccagg 5361  
DB 3 ATTCCAGTCACCTCTGTTGGTATCTATTACCTACAATGGGTCTCTCCAGG 62  
QY 5362 aaattgggatactgtaaagcccttctcctcagaggtggtgagatcctgcctactaa 5421  
DB 63 ATATTGGGATCTGTAAAGCCCTTCTCTCCAGAGGTGGTGTGCAGATCTGCTTACCAA 122  
QY 5422 actgtttgaagcaaaaaacacgtgtggtcaggtaccctagaaaaaagaaatagtttag 5481  
DB 123 ACTGTTTGAAGCAAAAAACACCGTGTGAGTACCTAGAAAAAGAAATAGTTGATAG 182  
QY 5482 acttctcgtgatactagctcctcctcctgaatacaagctctcattcagggtgccacggt 5541  
DB 183 AGCTTCTCTGATGACTATAGTGCCTCCTGAATCAAGCTTCTCATTTTCAGGTGCCACGGT 242  
QY 5542 ctgcagatgatgagaaagacatcctgtcctctgccttctctgtggtggtcctactatgtt 5601  
DB 243 CGCAGATGATGAGCAAGACATCTGTCTCTGCTCTGCTCTGCTGCTGCTGCTGCTGCT 302  
QY 5602 ctcaagaacattgtgcccaggaaattgtgaacgggggaagaggttgagcttgcatcttc 5661  
DB 303 CTCAGAACATCTGCTGCCAGGAAATGTGAACGGGGAAGAGGTTGGAGCTTGCAATTTTC 362  
QY 5662 acgcacttcaactgtgagccgagctgcatcttctcctaaaaatcagagaatgcgagtg 5721  
DB 363 ACGCACTTCACTGTGAGCGCGAGTGTGATTTTCTCTAAAAATCAGAGAATGCCGAGTGG 422  
QY 5722 tcttggttgaagtaaacgagcgtgctctcctcctcctcctcctcctcctcctcctcctcct 5781  
DB 423 TCTTGTTGAAGTAAGCCAGAGGCTGTGCTCTATCCAGCTCTCTTACTTGGATGAATATG 482  
QY 5782 gagaacagaccctggcctgaaagagggg-caaccccttcatttctcgtgagcgggtat 5840  
DB 483 GAGAACAGACCCCTGGCCTGAGAGGGGGCCACCCCTTCATTATCTGAGCGGTAT 542  
QY 5841 cggaagctccatttggtgtggaacaacactgcattatagaagagattgctagagccaa 5900  
DB 543 CGGAAGCTCCATTGTGTGTCGCAACAACACTGCATATTAGAAGAGGTGTGCTAGAGCCAA 602  
QY 5901 gggactaatcagatgtatttgattcaactggcagttactgtgagctcacaactcgcct 5960  
DB 603 GAGACTAATCAGATGTATTGTGATTCAACTGGCAGTTACTGTGAGCTCCAACTGCGCT 662  
QY 5961 caagacaatcacaatgacgacagtagtaagagctgattcaaaattatgaaacatttct 6020

Db 663 CAAGACAAATCAAAATGAGCAGTAGTAAGGCTGATTCCAAATATATGGAATCTTCT 722  
QY 6021 gagggctggaaagtattggagggtcttttgcctccatgccaggttcacttacatcaata 6080  
Db 723 GAGGCTGGTCAAGTATTGGAGGCTCTTTGCTCCATGTCAGGTTCACTTACATCAATA 782  
QY 6081 aaatattcttaatgga 6097  
Db 783 AAATATTCTTAATGGA 799  
RESULT 10  
LOCUS AK026998 2958 bp mRNA PRI 29-SEP-2000  
DEFINITION Homo sapiens cDNA: FLJ23345 fis, clone HEP13679.  
ACCESSION AK026998  
VERSION AK026998.1 GI:10439998  
KEYWORDS oligo capping; fis (full insert sequence).  
SOURCE Homo sapiens hepatoma cell\_line:HepG2 cDNA to mRNA, clone\_lib:HEP clone:HEP13679.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (sites)  
AUTHORS Kawabata, A., Hiki, T., Kobatake, N., Inagaki, H., Ikema, Y., Okamoto, S., Okitani, R., Ota, T., Suzuki, Y., Obayashi, M., Nishi, T., Shibahara, T., Tanaka, T., Nakamura, Y., Isogai, T. and Sugano, S.  
NEDO human cDNA sequencing project  
UNPUBLISHED (2000)  
REFERENCE 2 (bases 1 to 2958)  
AUTHORS Sugano, S., Suzuki, Y., Ota, T., Obayashi, M., Nishi, T., Isogai, T., Shibahara, T., Tanaka, T. and Nakamura, Y.  
Direct Submission  
SUBMITTED (29-AUG-2000) to the DDBJ/EMBL/GenBank databases. Sumio Sugano, Institute of Medical Science, University of Tokyo, Laboratory of Genome Structure Analysis, Human Genome Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan (E-mail: cdnal@ims.u-tokyo.ac.jp, Tel: 81-3-5449-5286, Fax: 81-3-5449-5416)  
COMMENT NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing; Research Association for Biotechnology; cDNA library construction, 5' - & 3' - end one pass sequencing; Department of Virology and Human Genome Center, Institute of Medical Science, University of Tokyo (partly supported by Science and Technology Agency).  
FEATURES  
Location/Qualifiers  
1..2958  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/cell\_line="HepG2"  
/cell\_type="hepatoma"  
/clone="HEP13679"  
/clone\_lib="HEP"  
/notes="cloning vector pME18SFL3"  
BASE COUNT 940 a 600 c 638 g 780 t  
ORIGIN  
Query Match 10.3%; Score 651.6; DB 9; Length 2958;  
Best Local Similarity 54.6%; Pred. No. 2.8e-153;  
Matches 1555; Conservative 0; Mismatches 1229; Indels 66; Gaps 10;  
QY 3113 agtggccacatttaagaacacaggtgtatcaggccatgagtttatgaactaaagatga 3172  
Db 9 AGTTGCCCATTTCAAGAAACCTGGATTAAACAGGAGGCGCATGTATGAATCAACACCA 68  
QY 3173 atcaactgaagacttcaatagtacttttctacttccaaacccagcagtagcaagc 3232  
Db 69 ATGTGCCAAAAGAGTTCACCTGTATTCTTCTATCTATCTATCTATCTATCTATCTATCT 128  
QY 3233 tgaacatatcagaagaagagagagaaacaaacaaagatgaagcattgcccacc 3292

1194 CCTTTGTAATGCTTGAATAAATACTGTTATTCC-----TCTGCTGCTTCTCTCCAAGAAA 1247  
QY 4349 aaacatgagaatgcagatgctctgtcacaatttttgacccttgccacggtggatagacag 4408  
1248 TATTTTAAACAACAGAGTTAAATTTTTCAGACCAACCAAAATCTGACTCAGTGGATTAGAAC 1307  
QY 4409 tgtctggccagaataatcagggttataataaagacatgctaaaggagagaacccaattcc 4468  
1308 AATATCTCAGCAATAAAGACATTACAGTTTCTTAGGAAAAGAAAGTAGTACTCCTAATAA 1367  
QY 4469 tatttcttaatacagaagatggagattctactctgttggagttccattccatccctgagttt 4528  
1368 TGCTCTACAAGAAATTCAGAAATGTGGATGAATTACACTCCCTGAAAGGTTGAGGCC 1427  
QY 4529 tggcgttgagttcgtattaaaattcaaatagcaatccaggcaatggtattctcttttc 4588  
1428 TGATTTTCTGCTTAAGATCCCTTATTTCTGAGAGCAATAAAGAAATGCTAACGACATTGG 1487  
QY 4589 caacaacatttagaattggattgaaagtgcacactgataaaggatcctcgagtcctc 4648  
1488 AACTGCTACTACAGGTGGGACTAAAGGTTTCATCCCAATGAAGAGGATCCTCGTGTTC 1547  
QY 4649 catgctgacctgagcacctgcgcttcaactatccaggcaatggaatctattggaga 4708  
1548 CATATGTGTGGGTAGCTGCGCGTACACCATCCAAAGCATAGAAAGAAATTTGAGTGA 1607  
QY 4709 tgaagaaaacacctctgtttggagcaactcaaaatagcagcataatggctgctgaagcatt 4768  
1608 TGAAGATAAACCATTCTTTGGTCTTTTACCTTGACAGCTGGATCAGTTTTCAGTGGTGAAGACATTT 1667  
QY 4769 aatgcagttgcaattgcacagagattacctgtcctcaggtcctgatacagaaaaatct 4828  
1668 GACGAGATTTGCGCGACACACTGGACAGTGGCATCAGTTTTCAGTGGTGAAGACATTT 1727  
QY 4829 ggtctgctctatcagttgcttctcaatacaatacaagatacacacatgcctct 4888  
1728 TTGTAAACCTTTTGGCATCACTGGTGCCTAATGACGCCATGAGGAATTCATGTCATATT 1787  
QY 4889 gtctatagatctgttctatgttttgggtgctgtgtgttagcattccatccctgtattg 4948  
1788 AGATATTGACATGTTTCATTTATTGGTGGCTTGGTGTGCTTGCATTTCTCGTGGCAGTG 1847  
QY 4949 ggaagacctgttgaatcgcagccttctcagttagttcttctcctataaacacctttatct 5008  
1848 -----TCAGGATTTTTCAGGGATCAGCTTGGCAGCTGGACCTGGACCTTCACAT 1892  
QY 5009 ctccatttgatcacatgacacatgctcagatactactitacagtagacagacgacct 5068  
1893 TTTCATCTGGTTACTATGGCACACATCATACAGATCTTACTTACCTCATGTACAGAAGA 1952  
QY 5069 acc---ccttgcacaggttcaagaagacagtggaagggtcattcccgcatctcttctct 5125  
1953 GAATGCGATGGATCAAGAAAATCCCCCTTGTGAAGAAGAAATCAGCAGTTCTTGTGTTGTA 2012  
QY 5126 tcagaaaattctcaatatacaagtggtccattgggtgtgatatctcctggctggtattt 5185  
2013 TAAACACTTTCACAGTATACGGGAAGTGCTTTGAAAAGAAATACCATCCGGCTGGCATCT 2072  
QY 5186 gtgggtcctcactgaagaatggcatccctctattctcgtctgctgcatgttttttcca 5245  
2073 GTGGAGAGGTGTGAGAGCTGGAATCATACCTTTCCCTGAAGTGTCTGCTTTATTTTTTCA 2132  
QY 5246 ctatttacttgggttaactcgcctgaggaactgcatlaccattctcagaagagagta 5305  
2133 TTAATTAATGGAGTTCTTCCCAACCCGACATTCACATCA---AGTTCTGGAACAACCCATTT 2189  
QY 5306 cadtgacctgttagctatctatttaccatacaaaatttctcgtctcctccaggaata 5365  
2190 TGAACATTTATGAGCTATCTTTCCCTACCAACAACCTCATTTGCTCTTTTCAAGAAA 2249  
QY 5366 ttgggactgttaagcccttgcctcagaggtgggtgtgagatcctccttactaaactg 5425  
2250 TAGTGAGATAATGAATTCAGTTGAAAGTTGTTGTCGCGTAACAGTGAAGTTAAAGATA 2309

1194 CCTTTGTAATGCTTGAATAAATACTGTTATTCC-----TCTGCTGCTTCTCTCCAAGAAA 1247  
QY 4349 aaacatgagaatgcagatgctctgtcacaatttttgacccttgccacggtggatagacag 4408  
1248 TATTTTAAACAACAGAGTTAAATTTTTCAGACCAACCAAAATCTGACTCAGTGGATTAGAAC 1307  
QY 4409 tgtctggccagaataatcagggttataataaagacatgctaaaggagagaacccaattcc 4468  
1308 AATATCTCAGCAATAAAGACATTACAGTTTCTTAGGAAAAGAAAGTAGTACTCCTAATAA 1367  
QY 4469 tatttcttaatacagaagatggagattctactctgttggagttccattccatccctgagttt 4528  
1368 TGCTCTACAAGAAATTCAGAAATGTGGATGAATTACACTCCCTGAAAGGTTGAGGCC 1427  
QY 4529 tggcgttgagttcgtattaaaattcaaatagcaatccaggcaatggtattctcttttc 4588  
1428 TGATTTTCTGCTTAAGATCCCTTATTTCTGAGAGCAATAAAGAAATGCTAACGACATTGG 1487  
QY 4589 caacaacatttagaattggattgaaagtgcacactgataaaggatcctcgagtcctc 4648  
1488 AACTGCTACTACAGGTGGGACTAAAGGTTTCATCCCAATGAAGAGGATCCTCGTGTTC 1547  
QY 4649 catgctgacctgagcacctgcgcttcaactatccaggcaatggaatctattggaga 4708  
1548 CATATGTGTGGGTAGCTGCGCGTACACCATCCAAAGCATAGAAAGAAATTTGAGTGA 1607  
QY 4709 tgaagaaaacacctctgtttggagcaactcaaaatagcagcataatggctgctgaagcatt 4768  
1608 TGAAGATAAACCATTCTTTGGTCTTTTACCTTGACAGCTGGATCAGTTTTCAGTGGTGAAGACATTT 1667  
QY 4769 aatgcagttgcaattgcacagagattacctgtcctcaggtcctgatacagaaaaatct 4828  
1668 GACGAGATTTGCGCGACACACTGGACAGTGGCATCAGTTTTCAGTGGTGAAGACATTT 1727  
QY 4829 ggtctgctctatcagttgcttctcaatacaatacaagatacacacatgcctct 4888  
1728 TTGTAAACCTTTTGGCATCACTGGTGCCTAATGACGCCATGAGGAATTCATGTCATATT 1787  
QY 4889 gtctatagatctgttctatgttttgggtgctgtgtgttagcattccatccctgtattg 4948  
1788 AGATATTGACATGTTTCATTTATTGGTGGCTTGGTGTGCTTGCATTTCTCGTGGCAGTG 1847  
QY 4949 ggaagacctgttgaatcgcagccttctcagttagttcttctcctataaacacctttatct 5008  
1848 -----TCAGGATTTTTCAGGGATCAGCTTGGCAGCTGGACCTGGACCTTCACAT 1892  
QY 5009 ctccatttgatcacatgacacatgctcagatactactitacagtagacagacgacct 5068  
1893 TTTCATCTGGTTACTATGGCACACATCATACAGATCTTACTTACCTCATGTACAGAAGA 1952  
QY 5069 acc---ccttgcacaggttcaagaagacagtggaagggtcattcccgcatctcttctct 5125  
1953 GAATGCGATGGATCAAGAAAATCCCCCTTGTGAAGAAGAAATCAGCAGTTCTTGTGTTGTA 2012  
QY 5126 tcagaaaattctcaatatacaagtggtccattgggtgtgatatctcctggctggtattt 5185  
2013 TAAACACTTTCACAGTATACGGGAAGTGCTTTGAAAAGAAATACCATCCGGCTGGCATCT 2072  
QY 5186 gtgggtcctcactgaagaatggcatccctctattctcgtctgctgcatgttttttcca 5245  
2073 GTGGAGAGGTGTGAGAGCTGGAATCATACCTTTCCCTGAAGTGTCTGCTTTATTTTTTCA 2132  
QY 5246 ctatttacttgggttaactcgcctgaggaactgcatlaccattctcagaagagagta 5305  
2133 TTAATTAATGGAGTTCTTCCCAACCCGACATTCACATCA---AGTTCTGGAACAACCCATTT 2189  
QY 5306 cadtgacctgttagctatctatttaccatacaaaatttctcgtctcctccaggaata 5365  
2190 TGAACATTTATGAGCTATCTTTCCCTACCAACAACCTCATTTGCTCTTTTCAAGAAA 2249  
QY 5366 ttgggactgttaagcccttgcctcagaggtgggtgtgagatcctccttactaaactg 5425  
2250 TAGTGAGATAATGAATTCAGTTGAAAGTTGTTGTCGCGTAACAGTGAAGTTAAAGATA 2309

QY	5426	tttgaagcaaaaaaacacccgtggtccaggtaccctagaaagaaataagtttgatagagct	5485
Db	2310	TCGAAAGGTGAAGAGAGATGCTATTAAGATATATCCAAGAGAAATCTAACAAATTAATAAACCCT	2369
QY	5486	tcttgatgactatagctgcctcctgaatacaagcttctcatttcagagtgccacgggtgc	5545
Db	2370	TCCAGAGGATATACACAGCCTCATTAATCAAGCATCCAATTTCTCGTCCCGCAATCAGG	2429
QY	5546	agatgatagcgaaagcatcctgctcctcgtgccttctgccttttctgtggggctatactatgttctca	5605
Db	2430	TGGTGATTAAGAGCAGAGGCCCAACTCTGTGCTTTGTGCGGATCTCTGCTGTGCTGCCA	2489
QY	5606	gaacatttgcgcagaaattgtgaacggggaagagttggagcttgccatttttccaagc	5665
Db	2490	GAGTTACTGCTGCCAGACTGAACCTGGAAGGGAGGATGTAGAGGCTGCACAGCTCACAC	2549
QY	5666	acttcacttgtggagcggagctgcctcatttctctaaaaatcagagaatccggagtggtct	5725
Db	2550	CTACTCCTGTGGCTGTGAGTGGGCATCTTCTTGAGACTACGGGAATGTCAGGTGCTATT	2609
QY	5726	ggttgaaggtaaaagcagaggtgctgcctatccagctccttacttggatgaatatggaga	5785
Db	2610	TTTAGCTGGCAAAACCAAGGCTGTTTTATTCTCTCCTTACCTTGATGACTATGGGGA	2669
QY	5786	acagacacctggcctgaagaggggcaaccccttcatattctcgtgagcgggtatcgga	5845
Db	2670	GACCGACCAGGAGCTCAGACGGGAAATCCTTTACATTTATGCAAGAGCGATTCAAGAA	2729
QY	5846	gtccatttgcctgcgaacaacactgcattatagaagagattgctgaagcccaagagac	5905
Db	2730	GATTGCAAGCTCTGGCACCAACAGTGTACAGAGGAAATTTGGCATGTCACAGGAAGC	2789
QY	5906	taatcagatgttattggattcaactggca	5935
Db	2790	CAATCAGACACTGGTGGCATTTGACTGGCA	2819
RESULT 11			
AC090514/c			
LOCUS	AC090514	166518 bp	DNA
DEFINITION	Homo sapiens chromosome 15 clone RP11-2016 map 15ql5, complete sequence.		PRI 01-MAR-2001
ACCESSION	AC090514		
VERSION	AC090514.1	GI:13129421	
KEYWORDS	HTG.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
AUTHORS	Rowen, L., Madan, A., Qin, S., Baradarani, L., Birditt, B., Bloom, S., Dors, M., Dickhoff, R., Fleetwood, P., Harrison, G., Kaur, A., Madan, A., Nesbitt, R., Traicoff, R. and Hood, L.		
TITLE	Sequencing of human chromosome 15 D15S146-D15S17 region		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 166518)		
AUTHORS	Rowen, L., Madan, A., Qin, S., Baradarani, L., Birditt, B., Bloom, S., Dors, M., Dickhoff, R., Fleetwood, P., Harrison, G., Kaur, A., Madan, A., Nesbitt, R., Traicoff, R. and Hood, L.		
TITLE	Direct Submission		
JOURNAL	Submitted (27-FEB-2001) Multimegabase Sequencing Center, Institute for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA 98105, USA		
REFERENCE	3 (bases 1 to 166518)		
AUTHORS	Rowen, L., Madan, A., Qin, S., Baradarani, L., Birditt, B., Bloom, S., Dors, M., Dickhoff, R., Fleetwood, P., Harrison, G., Kaur, A., Madan, A., Nesbitt, R., Traicoff, R. and Hood, L.		
TITLE	Direct Submission		
JOURNAL	Submitted (01-MAR-2001) Multimegabase Sequencing Center, Institute for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA 98105, USA		
COMMENT	----- Genome Center		

```

Center: Multimegabase Sequencing Center
Center code: UWMSC
Web site: http://chroma.mbt.washington.edu/msg_www
Contact: leerowen@systemsbiology.org
Drafting center: WIBR
----- Summary Statistics -----
Sequencing vector: pUC18; 108752
Chemistry: Dye-terminator Big Dye; 90% of reads
Chemistry: Dye-primer Big Dye; 10% of reads
Assembly program: Phrap; version 0.990399
-----
FEATURES             Location/Qualifiers
     source            1..166518
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /chromosome="15"
                        /map="15q15"
                        /clone="RP11-2016"
                        /clone_lib="RPCI human BAC library 11"
                        /note="This clone overlaps RP11-90M11 AC068727 and
RP11-473C18 AC068724"
     misc_feature      1..11505
                        /note="overlap with RP11-90M11 AC068727"
     unsure            47150..47190
                        /note="low quality data"
     unsure            106560..106790
                        /note="low quality data"
     unsure            109345..109360
                        /note="low quality data"
     unsure            114180..114240
                        /note="low quality data"
     misc_feature      156815..166518
                        /note="overlap with RP11-473C18, AC068724"
BASE COUNT      51718 a 34493 c 32351 g 47956 t
ORIGIN

```

[illegible]

```
QY 6282 cttctgcaatttggatgttgggtgtttt 6308
|||||
Db 41342 CTTCTGCCAATTTTGTATTTGGTGTTT 41316

RESULT 12
LOCUS AK026948 2512 bp mRNA PRI 29-SEP-2000
DEFINITION Homo sapiens cDNA: FLJ23295 fis, clone HEP10637.
ACCESSION AK026948
VERSION AK026948.1 GI:10439928
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens hepatoma cell_line:HepG2 cDNA to mRNA, clone_lib:HEP
clone:HEP10637.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (sites)
Kawabata,A., Hikiji,T., Kobatake,N., Inagaki,H., Ikema,Y.,
Okamoto,S., Okitani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T.,
Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.
NEDO human cDNA sequencing project
Unpublished (2000)
2 (bases 1 to 2512)
Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,
Shibahara,T., Tanaka,T. and Nakamura,Y.
Direct Submission
Submitted (29-AUG-2000) to the DDBJ/EMBL/GenBank databases. Sumio
Sugano, Institute of Medical Science, University of Tokyo,
Laboratory of Genome Structure Analysis, Human Genome Center;
Shirokane-dai, 4-6-1, Minato-Ku, Tokyo 108-8639, Japan
(E-mail:cdna@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,
Fax:81-3-5449-5416)
NEDO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan; cDNA full insert
sequencing: Research Association for Biotechnology; cDNA library
construction, 5'- & 3'-end one pass sequencing: Department of
Virology and Human Genome Center, Institute of Medical Science,
University of Tokyo (partly supported by Science and Technology
Agency).

FEATURES
source
1..2512
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="HepG2"
/cell_type="hepatoma"
/clone="HEP10637"
/clone_lib="HEP"
/note="cloning vector pME18SFL3"
BASE COUNT 785 a 511 c 543 g 673 t
ORIGIN

Query Match 7.7%; Score 487.8; DB 9; Length 2512;
Best Local Similarity 54.1%; Pred. No. 6.1e-112;
Matches 1220; Conservative 0; Mismatches 987; Indels 50; Gaps 9;

QY 3694 caggatcggaatctattagaatgatgagatctactcatgataaagaaagcagaacgaa 3753
|||||
Db 150 CAGAGCAGAGAACCCATAATGGAAGAGAGTTCAAGGGACAAAGACAAAGCTGAGAGGA 209
|||||
QY 3754 aaagaaagctgaagctgctaggtacatcgccagagaagatcatggctcagatgtctgcct 3813
|||||
Db 210 AGAGAAAGCAGAGATGGCCAGACTGCCGAGAGAAAGATCATGGCTCAGATGCTGAAA 269
|||||
QY 3814 tacagaaaaacttcattgaaactcataaactcatgtatgacatacatcagaatgcctg 3873
|||||
Db 270 TGCAGGGCATTTTATTGTATGAAAACAAAGAACTCTTTTCAGCAGACATTAGAACTGGATG 329
|||||
QY 3874 ggaagaagattccattatggaggaagagagcaccacagctcagctgactactctagaa 3933
|||||
Db 330 CCTCAACCTCTGCTGTTCTTGATCA-----TAGCCCTGTGGCTTCAGATATGACACTTA 383
|||||
```

```
Db 1438 TTACATTTTCCATCTGGTTACTATGACACATCATACATCTTACTTACCTCAGTA 1497
QY 5062 c-----aggcctacccttgcctcaggttcaagaagacagatgaagagcgtcattcccgcatctt 5118
Db 1498 CAGAAGAGAATGCGATGGATCAAGAAATCCCCCTTGTGAAGAAGAATCAGCAGTCTCTTG 1557
QY 5119 cttctttcagaaattctcaatatatacaagtgggtccctcattgggtgtagattctcggct 5178
Db 1558 CTTTGTATAAAACACTTACACAGTATACGGGAAGTGCCTTGAAGAATAACCATCCGGCT 1637
QY 5179 ggtatttgggtcctcactgaagaatggcatcaccccttatcttcgctgctgcattgt 5238
Db 1618 GGCATCTGTGGAGAGTGTCCAGAGTGGATCATGCCCTTCCCTGAAAGTGTCTCTTAT 1677
QY 5239 tttccactatttacttgggttaactcgcctgaggaactgcataccaattctgcagaag 5298
Db 1678 TTTTTCATTACTTAATGGAGTTCCTTCCCCACCCGACATCAAGT--TCCTGGAACAA 1734
QY 5299 gagagtacagtcactctgtagctatctattcttaccatacaaaattgtctctctcc 5358
Db 1735 GCAATTTGAACATTTATGTAGTATCTTCCCTTACCACAAACACCTCATTTGCCCTTTTC 1794
QY 5359 aggaatatgggatactgtaaagcccttgcctcagagtggtgagatcctgccttac 5418
Db 1795 AAGAAATAGTAGAATAATGAATCTACTGATGAAGTGTGGTCCGCTAAACAGTGAAGTTA 1854
QY 5419 taaactgtttgaagcaaaaaaacaccgtggtcaggttaccctagaaaaaagaaatgttga 5478
Db 1855 AAGATATCTAGAAGGTGAAGAGATGCTATAGATATCCAGAGATCTAACAAATAA 1914
QY 5479 tagagttcctgatctactatagctcctcctgaatcaagttctctcattcaggtgccac 5538
Db 1915 TAAACCTTCAGAGGATTAACAGAGGCTCATTAATCAAGCATCAATTTCTCGTGCCTGA 1974
QY 5539 ggtctcaatagatgacaaagcactcctcctcctctctctctctctctctctctctctat 5598
Db 1975 ATACAGTGTGTATAGAGCAGAGCCCAACTCTGTGCCCTGTGTGCGGATCTCTGTGT 2034
QY 5599 gttctcagaacattgtgcaggaattgtgaacggggaaggttggagcgttgctatt 5658
Db 2035 GCTCCAGAGTTACTGCTGCCACTCAACTGGAGGGGAGGATGTAGGAGCTGCACAG 2094
QY 5659 tbaacgacttcaactgtgagccgagctgtcatttctcctaaaaatcagagaatgccag 5718
Db 2095 CTCACACCTACTCCTGTGGCTCTGGAGTGGCATCTTCCCTGAGAGTACGGGAATGTCAGG 2154
QY 5719 tggctcgttgaaggtaaagccagaggtgtgcctatccagctccttacttggatgaat 5778
Db 2155 TGCTATTTTGTAGTGCAAAACCAAGGCTGTTTTTATTTCTCCTCCCTACCTTGATGACT 2214
QY 5779 atggagaacacagaccctggcctgaagaggggcaaccccttcttattctcgtgagcgt 5838
Db 2215 ATGGGAGACCCGACAGGACTCAGAGGGGAATCTTTTACATTTATGCAAGAGCGAT 2274
QY 5839 atcggaagctccatttggcttggcaacacactgcattatagaagagattgctaggagcc 5898
Db 2275 TCAAGAAGATTCAAGAGCTCTGGCACCACACAGAGTGTACAGAGGAATAATGGACATGCAC 2334
QY 5899 agagactaatacagatgttatttggattcaactggca 5935
Db 2335 AGGAAGCCAATCAGACACTGGTTGGCATTTGACATGGCA 2371

RESULT 13
AC068724/c
LOCUS
DEFINITION Homo sapiens chromosome 15 clone RP11-473C18 map 15q15, complete
sequence.
ACCESSION AC068724
VERSION AC068724.7 GI:12739735
KEYWORDS HTG.
SOURCE human.
```

Query Match

5.6%; Score 354; DB 9; Length 190727;

```
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 190727)
Rowen, L., Madan, A., Qin, S., Baradarani, L., Birditt, B., Bloom, S.,
Dors, M., Dickhoff, R., Fleetwood, P., Harrison, G., Kaur, A., Madan, A.,
Nesbitt, R., Traicoff, R. and Hood, L.
Sequencing of human chromosome 15 D15S146-D15S117 region
Unpublished
2 (bases 1 to 190727)
Rowen, L., Madan, A., Qin, S., Abbasi, N., Baradarani, L., Birditt, B.,
Bloom, S., Dors, M., Dickhoff, R., Fleetwood, P., Harrison, G., Kaur, A.,
Madan, A., Nesbitt, R., Traicoff, R., Shaffer, T. and Hood, L.
Direct Submission
Submitted (08-MAY-2000) Multimegabase Sequencing Center, Institute
for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA
98105, USA
3 (bases 1 to 190727)
Rowen, L., Madan, A., Qin, S., Baradarani, L., Birditt, B., Bloom, S.,
Dors, M., Dickhoff, R., Fleetwood, P., Harrison, G., Kaur, A., Madan, A.,
Nesbitt, R., Traicoff, R. and Hood, L.
Direct Submission
Submitted (10-FEB-2001) Multimegabase Sequencing Center, Institute
for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA
98105, USA
COMMENT On Feb 10, 2001 this sequence version replaced gi:11138174.
----- Genome Center
Center: Multimegabase Sequencing Center
Center code: UWMSC
Web site: http://chroma.mbt.washington.edu/msg_www
Contact: leerowensystemsbiology.org
----- Summary Statistics
Sequencing vector: pUC18; 108752
Chemistry: Dye-terminator Big Dye; 90% of reads
Chemistry: Dye-primer Big Dye; 10% of reads
Assembly program: Phrap; version 0.990399
Note: Data from overlapping BACs AC009825 [Drafting center: WIBR]
and AC009852 [Drafting center: UWMSC] were added for finishing
-----
FEATURES
Location/Qualifiers
1..190727
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="15"
/map="15q15"
/clone="RP11-473C18"
/clone_lib="RPCI human BAC library 11"
/note="This clone overlaps RP11-2016 AC021899, RP11-19D14
AC009825 and RP11-402F9 AC009852. Data from AC009825 and
AC009852 were added and the consensus sequence determined
from RP11-473C18 to the extent possible"
1..9704
/note="overlap with RP11-2016, AC021899"
42130..42140
/note="low quality data"
43640..43830
/note="low quality data"
69905..70000
/note="single read coverage"
87320..190727
/note="overlap with RP11-19D14, AC009825"
99880..99900
/note="low quality data"
103370..103380
/note="low quality data"
105748..190727
/note="overlap with RP11-402F9, AC009852"
BASE COUNT 54575 a 40945 c 40610 g 54597 t
ORIGIN
```

	Best Local Similarity	100.0%;	Pred.	No. 5.6e-78;	Matches	334;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
QY	423	atgcaacagttccctcaggaccacaacccccagatatcaaacatttcagtactaccaagagctacgc	482											
Dd	45931	ATGCAACAGTTCCTCAGCACCAACCAGATATACACCATTCAGTTCCACTGCCGTGGTGCGTGCCTTC	45772											
QY	483	ctgtataaatagaggggaiaaaaaaaatcttccagtcaccttaacgtcgtagcgttgcttc	542											
Dd	45771	CTGTATAAATTAGAGGGGAIAAAAAAAATCTCCAGTCCCTTCAGCTCGTAGCGTTCGCTTC	45712											
QY	543	cggagaacggcccaggaagcacactcctcgttgtctgtgtctgaacccaacgaactcaggggcgcgtc	602											
Dd	45711	CGGAGAAGCGGCCGGAAGCCACTCTCTCGAGTCTCGGTCAAACCCGACTTCAAGGGCGCGTC	45652											
QY	603	gtaaaagtgcgtccctctctctcagaccgccacaggtttccgctctgcctcctgcgcgg	662											
Dd	45651	GTAANAAGTGTGCTCCCCTGTCTCTCCGACCGGCCACAGGTTTTCCGCTTCCCTGTGCCGCGG	45592											
QY	663	ggctgcgaactgcaggcgctcagtttccctcaaagtgcggacgaggagctggagggtact	722											
Dd	45591	GGTCGGCAACTGCAGGCGTCAGTTTCCCTCAAGATGCGGACGAGGAGCTGGAGGTA	45532											
QY	723	gagaggatgaaatcatcgcggaggttatcccccagaccctcagctctgtgcgatc	776											
Dd	45531	GAGAGGATGAAAATCAGCGGGAGTTATCCCCAGACCCCTCAGCGCTCGGCATCT	45478											
RESULT 14														
ACO90992														
LOCUS														
DEFINITION														
ACO90992														
ACO90992.1 GI:13431042														
HTG; HTGS_PHASED.														
human.														
ORGANISM														
Homo sapiens														
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;														
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.														
REFERENCE														
AUTHORS	Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Bouckngalter,B., Brown,A., Camarata,J., Collampiano,A., Chang,J., Choepel,Y., Colangelo,M., Collins,S., Compymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goette,M., Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme.W., Iliev,I., Johnson,R., Jones,C., Karatas,A., Lakocque,K., Lamarez,R., Landers,T., Lehocaky,J., Levine.R., Liu,G., MacLean,C., MacDonald.P., Marquis.N., Matthews.C., McCarthy.M., McEwan,P., McKernan.K., MCPheeters.R., Meldrim.J., Meneus.L., Mihova.T., Mlenga.V., Murphy T., Naylor,J., Nguyen,C., Norbu,C., Norman.C.H., O'Connor.T., P.O'Donnell.P., O'Neil.D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara.V., Raymond,C., Retta,R., Rieback.M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti.M., Roy.A., Santos,R., Schauer.S., Schuback.R., Seaman,S., Sevry,P., Sougnez.C., Spencer.B., Stange-Thomann,N., Stojanovic,N., Strauss.N., Subramanian,A., Talamas,J., Tesfaye.S., Theodore.J., Travers,M., Travis.N., Trigilio,J., Vassiliev,H., Viel.R., Vo.A., Wilson,B., Wu,X., Wyman,D., Ye.W.J., Young.G., Zainoun,J., Zembeke.L., Zimmer,A. and Zody.M.													
TITLE	Submitted Submission													
JOURNAL	Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA													
COMMENT	All repeats were identified using RepeatMasker: Smit, A.F.A. & Green, P. (1996-1997) <a href="http://ftp.genome.washington.edu/RM/RepeatMasker.html">http://ftp.genome.washington.edu/RM/RepeatMasker.html</a> ----- Genome Center -----													





/db\_xref="taxon:10090"  
/cell\_type="embryonal stem"  
/dev\_stage="embryo"  
/chromosome="2"  
143..399  
/gene="Ubr1"  
/number=2  
BASE COUNT 273 a 150 c 168 g 319 t  
ORIGIN

Query Match 3.4%; Score 215.2; DB 10; Length 910;  
Best Local Similarity 89.2%; Pred. No. 3.6e-43;  
Matches 232; Conservative 0; Mismatches 28; Indels 0; Gaps 0;  
Qy 777 tgggtgggacagcaagttgatttttatactgctttcttgcatcatttggcacaattggtg 836  
Db 143 TGGTGGGATCAGCAAGTTGATTCTTCTACTGCTTTCTTACATCATTGGCACAATTAGTG 202  
Qy 837 ccagaaatttacttctgctgaaatggaccagacttggaaaagcaggaggaagtgtaaa 896  
Db 203 CCAGAAATTTATTGCTGAGATGGACCCAGATTTGGAAAAGCAAGAGAGAGTGACAG 262  
Qy 897 atgtcaatattcactccactggaatggatggtacttatttggagaagatccagatatattgctta 956  
Db 263 ATGTCAATACTCACCTCTTGGAGTGGTACTTATTGGAGAGGATCCGGATATTGCTTTA 322  
Qy 957 gagaatttgaagcacagtgagcatttcagcttctgtggagggttttcaaaagtgagag 1016  
Db 323 GAGAAATTAACACAGTGGACCGTCCAGTGTGTGGGAGGTTTTTCAAAAGTGGAGAA 382  
Qy 1017 acaacctattcttgcaggga 1036  
Db 383 ACAACATATTCCTGTAGGTA 402

Search completed: March 1, 2002, 16:16:46  
Job time: 13031 sec

GenCore version 4.5  
Copyright (C) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 1, 2002, 13:51:46 ; Search time 261.48 Seconds  
(without alignments)  
20682.299 Million cell updates/sec.

Title: US-09-724-126A-1  
Perfect score: 6308  
Sequence: 1 gcaagaattgcacgagg.....aatttgattgtggtttt 6308

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_1101.\*  
1: /SID52/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /SID52/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /SID52/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /SID52/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /SID52/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /SID52/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /SID52/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /SID52/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /SID52/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /SID52/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /SID52/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /SID52/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /SID52/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /SID52/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /SID52/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /SID52/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /SID52/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /SID52/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /SID52/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /SID52/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /SID52/gcgdata/geneseq/geneseq/NA2000.DAT.\*  
22: /SID52/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	4435.4	70.3	6395	20 AAV99308	CDNA encoding mous
2	4435.4	70.3	6395	22 AAC86933	Nucleotide sequenc
3	2546.8	40.4	2550	22 AAH14878	Human CDNA sequenc
4	996.2	15.8	1001	22 AAX03300	Partial CDNA encod
5	996.2	15.8	1001	22 AAC86934	Nucleotide sequenc
6	813.6	12.9	3327	22 AAX35730	CDNA encoding a pr
7	813.6	12.9	3502	20 AAX35731	CDNA encoding a pr
8	656.8	10.4	733	21 AAA02411	Human colon cancer
9	647	10.3	712	22 AAH07621	Human CDNA clone (
10	638.6	10.1	756	21 AAA02327	Human colon cancer
11	246.8	3.9	807	20 AAX39891	Gastric cancer ass

12	241.6	3.8	264	21	AAA00233	Human colon cancer
13	191.6	3.0	800	20	AAV99892	Gastric cancer ass
14	159	2.5	455	20	AAV89674	EST clone DA490.
15	118	1.9	936	22	AAF58252	Oligonucleotide D1
16	118	1.9	936	22	AAF58254	Oligonucleotide D1
17	118	1.9	936	22	AAF58257	Oligonucleotide D1
18	118	1.9	936	22	AAF58259	Oligonucleotide D2
19	118	1.9	936	22	AAF58262	Oligonucleotide D2
20	118	1.9	936	22	AAF58265	Oligonucleotide D1
21	116.6	1.8	308	16	AAT1876	Human gene signatu
22	116.4	1.8	936	22	AAF58252	Oligonucleotide D1
23	116.4	1.8	936	22	AAF58254	Oligonucleotide D1
24	116.4	1.8	936	22	AAF58257	Oligonucleotide D1
25	116.4	1.8	936	22	AAF58259	Oligonucleotide D2
26	116.4	1.8	936	22	AAF58262	Oligonucleotide D2
27	116.4	1.8	936	22	AAF58265	Oligonucleotide D1
28	105.6	1.7	366	21	AAC00938	Human secreted pro
29	105.6	1.7	456	21	AAC00940	Human secreted pro
30	105.6	1.7	1461	21	AAC78096	Human cancer assoc
31	103.2	1.6	605	21	AAC00939	Human secreted pro
32	100.8	1.6	602	21	AAC10733	Human secreted pro
33	100.8	1.6	692	21	AAC10734	Human secreted pro
34	100.8	1.6	842	21	AAC10872	Human secreted pro
35	54.8	0.9	2793	21	AAAT0146	Plasmodium falcipa
36	52.4	0.8	244	22	AAF58238	Oligonucleotide D1
37	51.6	0.8	244	22	AAF58238	Oligonucleotide D1
38	49.2	0.8	1727	22	AAH17089	Human CDNA sequenc
39	48.6	0.8	4590	22	AAH24065	Yeast AOP9604-asso
40	46.2	0.7	581	22	AAH09232	Human CDNA clone (
41	45.2	0.7	11722	19	AAV34455	Human MHC class I
42	45	0.7	143068	21	AAF21105	Human low adenosin
43	45	0.7	143068	21	AAF21105	Human low adenosin
44	45	0.7	143068	21	AAA34983	Human adenosine re
45	45	0.7	143068	21	AAA35150	Human adenosine re

ALIGNMENTS

RESULT 1  
AAV99308  
ID AAV99308 standard; cDNA; 6395 BP.  
XX AAV99308;  
XX  
DT 25-MAR-1999 (first entry)  
XX  
DE CDNA encoding mouse a ubiquitin-protein ligase, Ubtl.  
KW Ubiquitin-protein ligase; Ubtl; mouse; ubiquitinylation; degradation;  
KW N-end rule pathway; stress-related muscle wasting; inhibitor; screen; ss.  
XX  
OS Mus sp.  
XX  
FH Key Location/Qualifiers  
CDS 115..5388  
FT /\*tag= a  
FT /product= Ubtl  
XX  
XX  
XX US5861312-A.  
XX  
XX  
XX 19-JAN-1999.  
XX  
XX  
XX 02-DEC-1997; 97US-0982956.  
XX  
XX 02-DEC-1997; 97US-0982956.  
XX  
XX (CALY ) CALIFORNIA INST OF TECHNOLOGY.  
XX  
XX Kwon YT, Varshavsky A;  
XX  
XX WPI; 1999-130395/11.  
XX  
XX P-PSDB; AAW84351.



Db 1868 tagtgcattgtgcggtcatagctctggaacccaaaatcctcaaaagtgtctgagaccttg 1927  
QY taagcatatctgcacactctctagacaccccttgctggctcttcactgtacgttttaagcaggc 2568  
Db 1928 taagcatatctgcacactctctagacaccccttgctggctcttcactgtacgttttaagcaggc 1987  
QY tgggtgctgtttcaagactgcagtaattgtgctcttttgaggactttcaagtagaggtac 2628  
Db 1988 taggtgctattcaagaactgcagtaattgtgctcttttgacagctttcaagtagaggtcc 2047  
QY tagtgcgaatatcctttacgtgtgctgtggtgtggttgcacaggtgtgtgctgagatgtggc 2698  
Db 2048 tagtgcgagtcaccgcgtgcgtgctgctggtctggtgctgaggtgtgtgctgagatgtggc 2107  
QY gaagaaatggactgtctcttattagccaggtgttttattaccagaagtgttcaagtgcagag 2748  
Db 2108 gaagaaacgggctcactcatcagccaggtttcttattatcaagatgtttaaagtgcagg 2167  
QY aagaaatgatgataaagatatcatcatccttcagatgtgtgcattgtgtcatttaagtgtccca 2808  
Db 2168 aggaatgtcacgaagaagatatcatcatccttcagatgtgagcatctataatggatccca 2227  
QY ataagttctgttactggttacttcagaggtatgaacttgcgaggtttttaaacaagacca 2868  
Db 2228 acaagttctgttactggttacttcagagatatgaacttactgtatgtcttttaaacaagacca 2287  
QY tatctacaaaagaccaggatttgattaaacaataatacatcaataagaagaatgtctc 2928  
Db 2288 tatccaaaagaccaggatttgattaaacagtataatacatataagaagaatgtctc 2347  
QY aggtctcatctatatgtgggtgagcgttatgtactggaggtgggaatgtgaccaaaag 2988  
Db 2348 aggtctcatctatatgtgggagaacgttatgtactggaggtgggaatgttaccagag 2407  
QY aagaggtcaaatgagagaataattcactctgttctgtcattgaaccatgccacacagt 3048  
Db 2408 aggaagttaatatgagagagattacttacttctgtcattgagccatgcccacagtg 2467  
QY ccattgcccataatttacctgagaatgaaataatgaaactggttagagaatgtcataa 3108  
Db 2468 ccattgcccagaacctacctgagaacgaaataatgaaactggttagagaatgtcataa 2527  
QY acaaagtggccacatttaagaacacaggtgtatcagggccatgagttatgaactaaaaag 3168  
Db 2528 acaaagtggccacatttaagaacacaggtgtcggggccatgagttatgaattgaaag 2587  
QY atgaatcactgaaagacttcaatatgtacttttatctattactccaaacccacatagca 3228  
Db 2588 atgaatcactgaaagacttcaatatgtacttttatctattactccaaacacacatagca 2647  
QY aggtgaaacatatgcagaagaaaagagaaaacaaagaaagatgaagcattgcccgc 3288  
Db 2648 aggtgaaacatatgcagaagaaaagagaaaacaaagaaagatgaagcattgcccgc 2707  
QY caccacccactcctgaattctgocctgtcttccagcaaaagtattacaccttctcaactgtg 3348  
Db 2708 cgcacctcctcagagttctgctgctgttccagcaaaagttagtccaacctgctcagctgtg 2767  
QY atatcatgatgtacattctcagagaccogtatttgagcggccaatagacacagattctaaact 3408  
Db 2768 atgttatgataacatctccagaccatctttgagcggcagtggaacggaggtcttaact 2827  
QY tgtgaccgaaggtgctcccaatggcttttcataattctggcattggctgggttttactagaag 3468  
Db 2828 tgcgacagaagggatgtcgcagatggcgtcccatataatggcactgggctgctggaag 2887  
QY agaaacacagcttcaaaagctcctggaagaagaatgaacatttgcatttttatacaagg 3528  
Db 2888 agaagcagcagcttcagaagctcctggaaggaaggtggcttttgacttttaccataaag 2947  
QY cttcaagattgggaaggttcagccat-----gaatacaaaatgcttttggaaaaac 3579  
Db 2948 cttcaagattgggaaggttcagccatgaatgctcagaataatacaaaatgctcttggaaagc 3007

QY tcaaagaaattcccagttagaagccagaagcatataacgttgatacttcaagatg 3639  
Db 3008 tcaaaggaatccccaattagaagccagaagcatataacgttgatacttccagatg 3067  
QY ttgacacagtgaagcgttaagagaaaaatctgttttaattgtagcaaccacatcaggt 3699  
Db 3068 ttgacacagtgaagcgttaagagaaaaatctgttttaattgtggaacacacttcaaggac 3127  
QY cggaattctattaaagaatgatgagattactcatgataaagaaaaagcagaaacagaa 3759  
Db 3128 tggagtgcattaagagtgagagattactcatgataaagaaaaagcagaaacagaa 3187  
QY aagctgaagctgtagggtacatcgcaggaagatcatggctcagatgtctgcttacaga 3819  
Db 3188 aagctgagggcgtaggcttcacgtccgaagatcatggccagatgtctgcttacaga 3247  
QY aaaacttcattgaactcataaactcatgtatgacaatacatcagaaaaatgctgggaag 3879  
Db 3248 aaaacttcattgaacccacaacactcatgtatgaatacgtcagaagtaacaggaag 3307  
QY aagattccattatgaggaagagagacccacagcagtcagtcactactagaatgctt 3939  
Db 3308 aagactccattatgaggaagagagacacactcagcagtcagtcagtcactagaattgctc 3367  
QY tgggtcctaaacggggtccatctgttactgaaaaagaggtgctgactgctaccccttggc 3999  
Db 3368 tgggcccctaaacggggtgctgttaccgaaaaagaggtgctgactgctacccctgccc 3427  
QY aagaagaacagagaggtgaaaaatagaaaaataatgcattatgcggcctgtgtccaga 4059  
Db 3428 aagaagaacagaggtgaaaaactagaaaaataatgcattatgttcagcatgtgtcaga 3487  
QY aatctactgcttaaccagcacaggggaaaaacccatagaactctcagagagacccctag 4119  
Db 3488 aatccacgcgcctaacccagcacaggggaagcctgctggaccacttaggggaaacactgg 3547  
QY accacttttcatggatccagacttggcatatggaacttatcacaggaagctgtggtcatg 4179  
Db 3548 acccttttcatggatccagacttggcatatggaacttatcacaggaagctgtggtcatg 3607  
QY taatgcagcagtgctgtggcagagatatttgaagcgtgtacagctgagctctcagcagc 4239  
Db 3608 taatgcagcagtgctgtggcagagatatttgaagcgtgtgagctgagctctcagcagc 3667  
QY qcattcgtgtgacttttgcactggaagtggaagatatcttgcctctttgcaaat 4299  
Db 3668 gcatcagctagacctgttgacctggagcggagcgtacactatgcccgtcctgcaagt 3727  
QY ctctgtgcaatctgtgtatccccatttcttggcaacctcaaaagataaaacagtggaga 4359  
Db 3728 ctctgtgcaacactgtcatccccatccctttgacgcgcgagagaatcaacagtggaga 3787  
QY atgcagatgctctgtctcaacttttgacctggcagcgtggatagacagctgttctggcca 4419  
Db 3788 atgcgagggctctgtctcaacttttgacctggcgcggtggatagacagctgttctggcca 3847  
QY gaatatcaggttataataagaacatgctaaagagaaaaac---caattcctattttct 4476  
Db 3848 gaatatcaggttataataagaacatgctaaagagaaaaac---caattcctattttct 3907  
QY ttaatcaagaagatggagagattctacttggagttccattccatcctgagtttggcgttg 4536  
Db 3908 ttaatcaagaagatggggttccacttttgagtttcaatccatcctgagtttggagttc 3967  
QY agtctcgattataattcaaatagcatcaaggaatgtttattctcttggccacaaca 4596  
Db 3968 agtctcggtgaaataattcaaatagtatcaaggaatgttcattctcttcgcacaaca 4027  
QY tttatagaattggattgaaagtcacactgtatgaagggatctctcgatcccatctgctga 4656  
Db 4028 tttacagaattggcctgaaagtgcctcctgtatgaactagacccacaggtgcccatagta 4087





QY 2329 ttttactcatgtttccaagagtggtgcttggtgatgaagaactcttacttctggtcttata 2388  
DB 1748 ttttgcctatgtttccaagagtggtgcttggtgatgaagaactcttactggtggtcttata 1807  
QY 2389 aagaatgtcacaaagctgtgatgagtgagtgaccaggtttcatatctagtacgaagacag 2448  
DB 1808 aagaatgtcacaaagctgttaagtggcgcagttacaatttcatgtccagcaccagaagacag 1867  
QY 2449 tagtacaatctgttgacatagtttggaaacaaagtcctcacagagtatctcgagatcttg 2508  
DB 1868 tagtgaattgtgcgtcatagtcctggaaacaaatccttacaagtgctcgagacccttg 1927  
QY 2509 taagctacaatctgcacactctctagaccctctgtagaccctctggtctctatgttacgtttaagaagcc 2568  
DB 1928 taagctacaatctgcacactctctagaccctctgtagaccctctggtctctatgttacgtttaagaagcc 1987  
QY 2569 tgggtgctgtttcaagacgtcatgaatttggcttcttcttggaggtcttcaagtagaggtac 2628  
DB 1988 taggtgtctattcaagactgcatagtatttggcttcttggaggtcttcaagtagaggtcc 2047  
QY 2629 tagtgaatctcttctacgttctggtgttggcttgcacaggttggctgagatgtggc 2688  
DB 2048 tgggtgagtagccgcgtgcctgctggtgctggtggtcaggttggctgagatgtggc 2107  
QY 2689 gaagaatggactgtctcttatttagccaggtgttttattaccagatgtttaagtgcagag 2748  
DB 2108 gaagaacgggtctcactcactcatcagccaggtttcttattatcaagatgtttaagtgcagag 2167  
QY 2749 aagaatgtatgaataagatatcatcatgcttcagatgttggtgatctttaaagtagccca 2808  
DB 2168 aggaatgtacgataaagatatcatcatgcttcagatgttggtgatctttaaagtagccca 2227  
QY 2809 ataagttctgttactggttacttcagaggtatgaacttgcagaggtttttaacaagacca 2868  
DB 2228 acaagttctgttactcgggtacttcagaggtatgaacttactgtactgtctttaaacaagacca 2287  
QY 2869 tatctcaaaagaccaggtatttgattaaacaataatacatcaataagaagaatgtcttc 2928  
DB 2288 tatccacaaagaccaggtatttgattaaacagataatacatataagaagaatgtcttc 2347  
QY 2929 aggtcctcatctatatttgggtgagcgttatgtacctgagtgagtggaatgtgaccaagag 2988  
DB 2348 aggtcctcatctatatttggggaacgttatgtacctgagtgagtggaatgttaccagag 2407  
QY 2989 aagagttcaaatgagaaatcattcaactgtcttgcttgcttgaccatgccacacagt 3048  
DB 2408 aggaggttaatagtagagagattactcaacttacttctgtcattgagcccatgccacacagt 2457  
QY 3049 ccattgcccacaaatttacctgagaatgaataataatgaactggcttagaagaatgtcataa 3108  
DB 2468 ccattgcccacaaacctacctgagaacgaataataatgaactggcttagaagaatgtcataa 2527  
QY 3109 acaaatggtccacatttaagaacaccaggtgtgatacaggccatgagttatgaactaaag 3168  
DB 2528 acaaatggtccacatttaagaacaccaggtgtgagggccatgagttatgaattgaag 2587  
QY 3169 atgaatcaactgaagacttcaatgatgtacttcttattacttcaaaacccagcatagca 3228  
DB 2588 atgaatcaactgaagacttcaatgatgtacttcttaccatttcaaaacacagcatagca 2647  
QY 3229 aggtgtgaacatgtcagaagaagaagagaaaaacaaagaacaaagatgaagcattgccgc 3288  
DB 2648 aggtgtgaacatgtcagaagaagaagagaaaaacaaagaacaaagatgaagcattgccgc 2707  
QY 3289 caccacacactcctgaattctgccttgccttccagcaaaagtattaaactcttcaactgtg 3348  
DB 2708 cgccactcctcagagttctgccttgccttccagcaaaagtattcaactcctcagcgtg 2767  
QY 3349 atataatgatgtacattctcaggaacaggtattttagaggggcaatagacacagattctcaact 3408  
DB 2768 atgttatgatatacctcctcaggaacatctttagaggggagtgaggggagtgagcaggtctcaatc 2827

QY 3409 tgtgagaccgaaggagtgctccaaatgggttttccattattctggcattgggtttactagaag 3468  
DB 2828 tgtgagacgaaggagtgctgcagatggcgttccattattctggcattgggtttactagaag 2887  
QY 3469 aagagcaacagcttcaaaaagctctcgaagaagtaacatttgacttttactataag 3528  
DB 2888 aagagcagcagcttcaaaaagctctcgaagaagtaacatttgacttttactataag 2947  
QY 3529 cttcaagaattgggaagttcagccat-----gaatatacaaatgcttttggaaaaaac 3579  
DB 2948 cttcaagaattgggaagttcagccatgaatgctcagaataatacaaatgcttttggaaaaaac 3007  
QY 3580 tcaagaagaattcccccagtagaagccagaggaacatataaacgttgatacttcagatgt 3639  
DB 3008 tcaagaagaattcccccagtagaagccagaggaacatataaacgttgatacttcagatgt 3067  
QY 3640 ttgacacagtgaaagcattgaagaagaatacttctttaaattgttagcaaccacatcagat 3699  
DB 3068 ttgacacagtgaaagcattgaagaagaatacttctttaaattgttagcaaccacatcagat 3127  
QY 3700 cggaaattcttaagaatgatgagattactcatgataaagaataaacgagaaacagaaagaa 3759  
DB 3128 tggagtgatttaagaagtgagagattactcatgataaagaataaacgagaaacagaaagaa 3187  
QY 3760 aagctgaagctgtaggtacatgcgcaagaagatcatggtcagatgtctgcttacaga 3819  
DB 3188 aagctgagccgctaggtctcatgcgcaagaagatcatggtcagatgtctgcttacaga 3247  
QY 3820 aaaacttcaattgaactcataaaactcatgtatgacaaatacatcagaataatgctgggaaaag 3879  
DB 3248 aaaacttcaattgaactcataaaactcatgtatgataacacgtcagaagtaacagggaggg 3307  
QY 3880 aagattccattatgaggaagagagcaccacagcagtcagtgacttacttagaattgctt 3939  
DB 3308 aagattccattatgaggaagagagcaccacagcagtcagtgacttacttagaattgctt 3367  
QY 3940 tgggtccttaaacgggtccatctgttactgaaaagaggtgctgcagtcagtccttctgccc 3999  
DB 3368 tgggccttaaacgggtccggtgtgtacgaaaagaggtgctgcagtcagtccttctgccc 3427  
QY 4000 aagaagaacagaggtgaaaatagaaaaataatgcatggattatcgccctgtgtccaga 4059  
DB 3428 aagaagaacagaggtgaaaatagaaaaataatgcatggattatgtcagcatgtgtcaga 3487  
QY 4060 aatctactgcttaaccacgacacaggggaaaccccatagaactctcagaggaagccctag 4119  
DB 3488 aatccacggccttaaccacgacacaggggaaagcctgtgtggaccacttaggggaaaacactgg 3547  
QY 4120 accacatttctcatggatccagacttggcatatggaaacttatcacagaagctgtggtcatg 4179  
DB 3548 acccttcttcatggatccagacttggcatatggaaacttatcacagaagctgtggtcatg 3607  
QY 4180 taatgcaacgagtgctgtgagagaagatatttgaagctgtacagctgagctcctcagcagc 4239  
DB 3608 taatgcaacgagtgctgtgagagaagatatttgaagctgtgagctgagctcctcagcagc 3667  
QY 4240 gcaatcattgtgacacttttgaacttgaagtgagaatacttctgcccctcttttcaaat 4299  
DB 3668 gcaatcattgtgacacttgaacttgaagtgagaatacttctgcccctcttttcaaat 3727  
QY 4300 ctctgtgcaatcactgtgaccccatattcttctgcaacctcaaaagaataacagtgaga 4359  
DB 3728 ctctgtgcaacactgtgaccccatcctcttgcagccgagagaagatcaacagtgaga 3787  
QY 4360 atgcagatgtcttctgctcaacttttgacctggcagctgtggaatacagacttcttggcca 4419  
DB 3788 atgcagaggtcttctgctcaacttttgacctggcagctgtggaatacagacttcttggcca 3847  
QY 4420 gaatatcaggttataataataagacatgtataaggaaggaacaccc---caattcctatttct 4476  
DB 3848 gaatatcaggttataataataagacatgtataaggaaggaacacccagcttcttctgt 3907  
QY 4477 ttaataaggaatgggagattctacttcttggaggttccattcctcctcctgagtttggcgtg 4536



||||| 3908 ttaatacgaagaatggggattcaacttttgagtttcaattccatccctgaattttggagttc 3967  
QY 4537 agtcttcgattaaatattcaaatagcatcaaggaaatgggtttattctctttcccaacaa 4596  
Db 3968 agtcttcggtgaatattcaaatagtatcaaggaaatgggtctctcttcgccacaa 4027  
QY 4597 ttatagaattggattgaagtgccactgatgaaggatccctcgagtcgccatgctga 4656  
Db 4028 ttacagaattggctgaagtgctctgatgaaccacgagtgcccatgga 4087  
QY 4657 cctggagcactgccttcaactatcaggaactgaaatctattggagatgaagaa 4716  
Db 4088 cctggagcagtgctgtcaactccaggcaatcgaaaaactgtgggagatgaagaa 4147  
QY 4717 aacctgtttggagcacttcaaataggcagcataatgggtctgaaagcattaatgcagt 4776  
Db 4148 aacctgtattggagcacttcaaatagacagcatagcgtctgaaggcgtaatgcagt 4207  
QY 4777 ttgcagttgcacagagattacctgtcctcaggtcctgatcacagaacatctgttcgtc 4836  
Db 4208 ttgcagttgcacagaggtacctgtcctcaggtcctgatcacagaacatctgttcgtc 4267  
QY 4837 ttctacagttgtcttccaaataaaatcagaagatcacaccatgcctctgtctatag 4896  
Db 4268 tctgtcagttattcttccaaactgcaatcagaaaaatacacccggtctctgtctg 4327  
QY 4897 atctgttttcattgtttgggtgctgtgttagcattcccatcctgttatgggatgacc 4956  
Db 4328 atcttccatgttctggtgcgcagctcttaggttcccatcctgttatgggatgaca 4387  
QY 4957 ctgtgtatcgcagcctcttcagtttagttcttccataaacacacttatctctccatt 5016  
Db 4388 ccgtggatctgcagcgtgcacacttagttctcatataaacacactctatcttccatc 4447  
QY 5017 tgatcacatggcaacatgcttcagataactactacagtagaca-----cag 5064  
Db 4448 tgatcacatggcgacatgcttcagataactccttaacacagatcacagatgtgtccag 4507  
QY 5065 gcttacccctgtcaggttcaagaagacagtggaaggctcattccgcatctcttctt 5124  
Db 4508 ggcgcgcgttctgtgaggtgaagagtagtgaggaggtcgtgtgcatctgcttct 4567  
QY 5125 ttgcagaaattctcaatatacagaagtgctccattgggtgtgatatctcgtgctgatt 5184  
Db 4568 ttgtggaagtgtgcagcacacagcgcctcactgggtgcggtgcctccggtgctacc 4627  
QY 5185 ttgtggtctcaactgaagaatggcatcacccttatcttcgtgtgctgcatgttttcc 5244  
Db 4628 ttgtggtctcctcaggaaggcatcacccttaccctcctgtgtgctgactgttttcc 4687  
QY 5245 actatttacttgggtgaactccgctgaggaaactgcataccaattctgcagaaggagat 5304  
Db 4688 actatttacttgggtgagtcgcgcctgaagaaactgtgtgcaattctgctgaaggagat 4747  
QY 5305 acagtgcaactctgtagctatctattacatacaaaattgttctcgtctctccaggaa 5364  
Db 4748 tcagtgcaactctgtagctatctattaccacaaaattgttctcgtcttccaggaa 4807  
QY 5365 attgggatactgtgaagcccttgcacagaggtggtgtgcagatccctgcttactaaact 5424  
Db 4808 attgggataccataaggcccttactacagaggtggtgtgcagatccctgcttactcaagt 4867  
QY 5425 gtttgaagcaaaaaaacacogtgcaggttaccctagaaaaaagaatagttgtatagagc 5484  
Db 4868 ctttgaagcagaaaaagtcgtgtggtcaggtacccttagaaaaaagaatagttgtatagagc 4927  
QY 5485 ttctgtagtactatagtgctcctcgtgaatcaagttctcatttcaggtgccacaggtctg 5544  
Db 4928 ttctgtaggactacagctgtcttctaaatcagggtctctcacttaggtgtccacggtctg 4987  
QY 5545 cagatgatgacgaagaacatcctgtccttgccttttctgtgggctatactatgttctc 5604  
|||||

Db 4988 cagatgatgagcgaagaacatcctgtcctctgtctgttttctgtggggccatcctgtgttctc 5047  
QY 5605 agaacatttgcgtccaggaaattgtgaacggggaaagggttgagcttgcatttttccag 5664  
Db 5048 agaacatttgcgtccaggaaattgtgaatgggaagggttgagcgtggttttccatg 5107  
QY 5665 caattcacttggagcccgagctcgtcatttctctaaataatcagagaatgccagatgggtcc 5724  
Db 5108 cgtctcatttgtgtggtcgtgagctgcatcttctctaaataatccgagaatcgaggtgtcc 5167  
QY 5725 tggttgaaggttaaacgcagagcgtgtgcttatccagctccttacttggatgaatatggag 5784  
Db 5168 tggttgaaggaataaacgcagagcgtgtgctaccacgcccttacttggatgaatatggag 5227  
QY 5785 aaacgacacctggcctgaagaggggcaaccccttctattctctgtgagcgttatcgga 5844  
Db 5228 aaacgacacctggcctaaagagaggaaacccacttctattctcggagcgttatcgga 5287  
QY 5845 agtccatttggctggcaaacacactgcattatagaagagattgctagagcgaagaga 5904  
Db 5288 agtctcatttggctggcaacacactgcattatagaagagattgctcggagcgaagaga 5347  
QY 5905 ctaatcagatgttatttggattcaactggcagttactgtgagctccaactctgctcaag 5964  
Db 5348 ctaatcagatgtctatttggatttaactggcagttactgtgagcctcagttcctcctcaag 5407  
QY 5965 acaatcacaaatgcagcagtagtaaaaggctgattcaaaattatggaaaaacttctcgag 6024  
Db 5408 acaatcagtagtgacatcaataataaagactgatctaaaattctagagaacttctcgag 5467  
QY 6025 gctggaaagtattggaggggttttgcctcctgcaggttcacacttcacatcaataaaat 6084  
Db 5468 acgggggaagtattggaggggtcttttgatccatgtccagattcacacacattataaaat 5527  
QY 6085 atttcttaagttagattgtcttcaattagcaaacatagcttccacaggaaaaa-aggac 6143  
Db 5528 attctttaaagtaattgtcttcaattatcaacataagcttcaaggggaaaaacaagac 5587  
QY 6144 atagatcaaatctgt-----ttt 6160  
Db 5588 atagattaaatgttttatgttttagaacactaaagaatgcttttccatccaagtgtctat 5647  
QY 6161 atgtctagtagtttccaggaaatttattcccttcaataattgtctcatttcttattt 6220  
Db 5648 ttctgtaatttccagaaaaactccttcccttcaataactgctcctagttcattcattat 5707  
QY 6221 catccacttggtagatgaagtcacgtcacaacagttgtagacattttatgtgtgttaac 6280  
Db 5708 caccacactggttaatgaggtcacattgaagcatttggacatttctccatctggtcaac 5767  
QY 6281 tcttctgcaatttctgtgtgtttt 6307  
Db 5768 atctctgcaacttctgtatttgggtttt 5794

## RESULT 3

AAH14878

ID AAH14878 standard; cDNA; 2550 BP.

XX

AAH14878;

XX

26-JUN-2001 (first entry)

XX

Human cDNA sequence SEQ ID NO:12731.

DE

Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

XX

Homo sapiens.

OS

Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

XX

Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

XX

Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

XX

Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

XX

Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

```
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
XX (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
XX WPI: 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs.
XX
XX Claim 8; SEQ ID 12731; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 2550 BP; 830 A; 454 C; 558 G; 708 T; 0 other;

Query Match 40.4%; Score 2546.8; DB 22; Length 2550;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2548; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1189 atcatgaacctgaagacagcagctactataaagagaattcacgctgcccgttgaaatgaag 1248
DB 1 atcatgaacctgaagacagcagctactataaagagaattcacgctgcccgttgaaatgaag 60

QY 1249 aggttaattgccaagccaggaataatttccttcagtgataaaatgctcgtagaatga 1308
DB 61 aggttaattgccaagccaggaataatttccttcagtgataaaatgctcgtagaatga 120

QY 1309 ctatatgggaagagaaagaaactgctcctcgaactccagataaaggagaaataatga 1368
DB 121 ctatatgggaagagagaaagaaactgctcctcgaactccagataaaggagaaataatga 180

QY 1369 gatactattgtctctttcattgatgaacacattcatatgacacgctcatatagacc 1428
DB 181 gatactattgtctctttcattgatgaacacattcatatgacacgctcatatagacc 240

QY 1429 tacaaagagcctcttgaactgtgagctgcagagccagcttgatataccactgaccttgaca 1488
DB 241 tacaaagagcctcttgaactgtgagctgcagagccagcttgatataccactgaccttgaca 300

QY 1489 aagaggggtcgctgttaaagcgggagcgttatgtctgtccaggaagcaaaaggaag 1548
```



```
QY 2979 gtgacaaagaagaggtcacaaatgagagaataatcattcacttgcttgcattgaacccatg 3038
Db 181 gtgacaaagaagaggtcacaaatgagagaataatcattcacttgcttgcattgaacccatg 240
QY 3039 ccacacagtgcattgcacaaatctactgagaaatgaaaaaataatgaaactggcttagag 3098
Db 241 ccacacagtgcattgcacaaatctactgagaaatgaaaaaataatgaaactggcttagag 300
QY 3099 aatgtcataaacaagtgccacatttaagaaacacagtggtatcatcaggccatggagtttat 3158
Db 301 aatgtcataaacaagtgccacatttaagaaacacagtggtatcatcaggccatggagtttat 360
QY 3159 gaactaaagaatgaatcactgaagacacttcaatgtactcttttattcattactccaaaaacc 3218
Db 361 gaactaaagaatgaatcactgaagacacttcaatgtactcttttattcattactccaaaaacc 420
QY 3219 cagcatagcaggctgaacatagcagaagaaagagaaacacagaaacacaaagatgaa 3278
Db 421 cagcatagcaggctgaacatagcagaagaaagagaaacacagaaacacaaagatgaa 480
QY 3279 gcaattgcgcaccaccactcctgaattctgcctgtctttcagcaaaagtgattacacct 3338
Db 481 gcaattgcgcaccaccactcctgaattctgcctgtctttcagcaaaagtgattacacct 540
QY 3339 ctcaactgtgatatcatgatgtacattctcaggaacgttatttgagcgggcaatagacaca 3398
Db 541 ctcaactgtgatatcatgatgtacattctcaggaacgttatttgagcgggcaatagacaca 600
QY 3399 gattctaactgtgacgaagggatgctcaaatggcttttcatattctgcattgggt 3458
Db 601 gattctaactgtgacgaagggatgctcaaatggcttttcatattctgcattgggt 660
QY 3459 ttactagaagaagaacacagcttcaaaaagctcctgaagaagaagtaacatttgacttt 3518
Db 661 ttactagaagaagaacacagcttcaaaaagctcctgaagaagaagtaacatttgacttt 720
QY 3519 tatctaaggcttcaagattgggaagttcagccatgaatatacaaatgcttttgaaaaa 3578
Db 721 tatctaaggcttcaagattgggaagttcagccatgaatatacaaatgcttttgaaaaa 780
QY 3579 ctcaagaataccccaggttagaagccagagaagacatgataacgtgatacttcagatg 3638
Db 781 ctcaagaataccccaggttagaagccagagaagacatgataacgtgatacttcagatg 840
QY 3639 ttgacacagtgaaagcattgaagaagaaatcttgaattgtgacacaccatcagga 3698
Db 841 ttgacacagtgaaagcattgaagaagaaatcttgaattgtgacacaccatcagga 900
QY 3699 tcgggaattattagaatgatgagattactcatgataaagaagaaagcagaagaaaaaga 3758
Db 901 tcgggaattattagaatgatgagattactcatgataaagaagaaagcagaagaaaaaga 960
QY 3759 aaagctgaagctgtaggctacatgcgcagaagatcgtgc 3799
Db 961 aaagctgaagctgtaggctcattcgcagaagatcgtgc 1001
```

## RESULT 5

AAC86934

ID AAC86934 standard; cDNA; 1001 BP.

XX

AC

AAC86934;

XX

DT 02-APR-2001 (first entry)

XX

DE Nucleotide sequence of a partial human Ubrl protein.

XX

KW Ubrl; E3-type protein; ubiquitin system; ubiquitin-protein ligase;

KW N-end rule pathway; intracellular pathogen; Lysteria monocytogenes;

XX

KW Yersinia enterocolitica; muscle wasting; infection; ss.

XX

OS Homo sapiens.

XX

XX

XX

FH Key Location/Qualifiers  
FT CDS l..999  
FT /\*tag= a  
FT /product= "Ubrl"  
FT /note= "partial sequence"

XX US6159732-A.

XX 12-DEC-2000.

XX 11-JAN-1999; 99US-0228317.

XX 02-DEC-1997; 97US-0982956.

XX (CALY ) CALIFORNIA INST OF TECHNOLOGY.

XX Kwon YT, Varshavsky A;

XX WPI; 2001-090278/10.

XX P-PSDB; AAB31163.

XX Inhibiting the N-end rule pathway in mammalian cells for treating  
PT infections and various diseases associated with muscle tissue wasting,  
PT by inhibiting the expression of Ubrl gene

XX Claim 4; Column 27-30; 18pp; English.

XX The present sequence encodes a partial Ubrl enzyme. Ubrl is an E3-type  
CC protein of the ubiquitin system. Specifically, it is a ubiquitin-protein  
CC ligase. The enzyme is specific for destabilising residues exposed at  
CC the N-terminus of protein substrates. Inhibition of the expression of  
CC Ubrl gene in a cell results in inhibition of the N-end rule pathway.  
CC The method is used for treatment of mammalian cells infected with an  
CC intracellular pathogen, e.g. Lysteria monocytogenes or Yersinia  
CC enterocolitica. Inhibition of N-end rule pathway is also useful for  
CC treating various diseases associated with wasting of muscle tissue and  
CC infections.

XX Sequence 1001 BP; 363 A; 186 C; 205 G; 247 T; 0 other;

Query Match 15.8%; Score 996.2; DB 22; Length 1001;  
Best Local Similarity 99.7%; Pred. No. 1.9e-261;  
Matches 998; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2799 atggatcccaataagttctgttactgttactcagaggtatgaactgcgaggtcttt 2858

Db 1 atggatcccaataagttctgttactgttactcagaggtatgaactgcgaggtcttt 60

QY 2859 aacaagaccatattcacaagaaccaggtttgattaaacaataatacactaatagaa 2918

Db 61 aacaagaccatattcacaagaaccaggtttgattaaacaataatacactaatagaa 120

QY 2919 gaaatgcttcaggtcctcatctatttgggtgagcgttatgtacctggagtggaat 2978

Db 121 gaaatgcttcaggtcctcatctatttgggtgagcgttatgtacctggagtggaat 180

QY 2979 gtgacaaagaagaggtcacaaatgagagaataatcattcacttgcttgcattgaacccatg 3038

Db 181 gtgacaaagaagaggtcacaaatgagagaataatcattcacttgcttgcattgaacccatg 240

QY 3039 ccacacagtgccattgccccaaatatttacctgagaaatgaaaaatgaaactggcttagag 3098

Db 241 ccacacagtgccattgccccaaatatttacctgagaaatgaaaaatgaaactggcttagag 300

QY 3099 aatgtcataaacaagtgccacatttaagaacacaggtgtatcagggcattggagtttat 3158

Db 301 aatgtcataaacaagtgccacatttaagaacacaggtgtatcagggcattggagtttat 360

QY 3159 gaactaaagaatgaatcactgaaagacattcaatgtacttttattcattactccaaaaacc 3218

Db 361 gaactaaagaatgaatcactgaaagacattcaatgtacttttattcattactccaaaaacc 420



Db 668 aagatacagactccacactccggtgtgtgctccacttgcctctgtttgcaagcctgg 727  
QY 3331 ttaactcttcaactgtgatataatgatgtacattctcagaccgtattttgagcgccaa 3390  
Db 728 ttaacattttgagtcagatgcatgtgtgcatcatggaacaaattctgcaatggcgtg 787  
QY 3391 tagacacagattctaaactgttgagaccgaaggtatgctccaaatggcttttcaattcttg 3450  
Db 788 tggaaacataatggatgctgctggtcagagtcctatgctgcaaaagggtgttacatttaattg 847  
QY 3451 cattgggtttactagaagaagcaacagcttcaaaaagcctcctgaag---aagaagtaa 3507  
Db 848 gcatggcactacaagaagaaacacatttagagaatgtccggaagagcatgtagttaa 907  
QY 3508 catttgacttttatacagaaggtccaagattgggaagttcagccatgaat-----a 3558  
Db 908 catttaccttactcagaagatatcaaaacctgggaagcgccaaaaaattctcttagca 967  
QY 3559 tacaatgccttttgaaaaactcaagaagattcccccagttagaagccagagagcatga 3618  
Db 968 tactagctatgctggaacacactacaaaaatgctccctaacctagaagtcccaagaagcatga 1027  
QY 3619 taactgtgatactcagatgtttgacacagtgaaagcgttagagagaataatctgttttaa 3678  
Db 1028 ttcggtggtatattgaagacttttaatgctgttaaaagatgaggaga-----gttcac 1081  
QY 3679 ttgtgacaaccacatcaggatcggaatctattaaagaatgagattactcatgataaag 3738  
Db 1082 ctaccagtcctgtgcagagacagaagaaccataatggaagagagttcaaggggcaaaag 1141  
QY 3739 aaaaagcagaaagaaagctgaagctgtaggtctacatcgccagagaagatcatgg 3798  
Db 1142 acaaaagctgagaggaagaaagacagagattgccagactgcgacagaaagaagatcatgg 1201  
QY 3799 ctcaagatgctccttacagaaaaacttcattgaaactcataaaactcatgtatgacaata 3858  
Db 1202 ctcatgattgtgaatgacggcgtattttattgatgaacaaagaacctctttcagcaga 1261  
QY 3859 catcagaatgctgggaagaaagattccattatgaggaagagagacccacagcagtcga 3918  
Db 1262 cattggaactgtagctcaacctctgtgtgttcttgatca-----tagcctgtgtgctt 1315  
QY 3919 gtgaactactagaattgctttgggttcctaaacgggtccatctgttactgaaaaagagg 3978  
Db 1316 cagatatgacacttacagacactggggcccgcaaaactcaggttctctgaaacaaagacaat 1375  
QY 3979 tgcctgactgcatctcttggccaaagaacagagaggtgaaaaatagaaaaataatgcccattg 4038  
Db 1376 tcgtttacatgtatatgtgtcaagagggagcaagaagttaaagtggaaagcaggggcaatgg 1435  
QY 4039 tattatcgccgtgtgtccagaagaatactactgccttaaccacagacaggggaaaaaccatag 4098  
Db 1436 tcttggcagcatgtgttcagatgaacaactgtattatcaaaaacagaagaataatttatic 1495  
QY 4099 aactctcaggagaagcccttagaccacttttcatggtatccagacttggcatatggaaactt 4158  
Db 1496 aagatccagaaaaaatat---gatccattattcatgcacctgatctgtgttgtggaaacac 1552  
QY 4159 atacaggaagctgtgtgctatgtaatgcacgcaggtgtgctggcagaagatattttgaaagctg 4218  
Db 1553 acactagtagctgtgggcacattatgcatgcccattgttggcaagggtattttgattccg 1612  
QY 4219 tacagctgagctctcagacagcattcatgtgtgacactttt-----tgactttgg 4266  
Db 1613 ttcagctaaagaacaggaaggaacagagattacgccttacatcagcagctatgtatgtag 1672  
QY 4267 aaagtggagaatatcttggcctcttctgcaaatctctgtgcaactgtgaccccatla 4326  
Db 1673 aaaaaggagaaatctcttggccctcttggaaatgcttggatgcttgaatctgtattctctctgc 1732  
QY 4327 ttctcttgcaacctcaaaagataaaacagtgagaatgcatgctcttctgtcacttttga 4386

Db 1733 tgcttt-----ctccaagaataatttttaacaacaggtttaaaattttttcagaccacaa 1786  
QY 4387 ccttgccacggttgatagacactgttctgcccagaaatacaggtttataataagaacatg 4446  
Db 1787 atctgactcagtggttagaacaatactcagcaaaataaaagcattacagtttcttaga 1846  
QY 4447 ctaaaagagaaaaaccaaattcctattttttaaatacgaaggaatgggagatttctactttgg 4506  
Db 1847 aagaagaaagtactcctaaataatgcctctacaagaattcagaanaatgtggatgaattac 1906  
QY 4507 agttcattccactcctcaggttttggcgttgagttcctgcttaataatttcaaaatagacca 4566  
Db 1907 agtccctgaagggttcaggccctgattttcgtcctagatcccttattctgagagcataa 1966  
QY 4567 aggaatgtgttattctcttggccacaacaatttatagaattgggattgaaagtgcacactg 4626  
Db 1967 aagaatgtctaaagacatttggaaactgtacctacaaggtgggactaaaagttcatocca 2026  
QY 4627 atgaaggggatacctcagagtcctcagctgagccctggagacacctgcgcttctaactatccagg 4686  
Db 2027 atgaagaggatcctcgtgttcccataatgtgtgggtagctgcgctacaccatccaaa 2086  
QY 4687 caattaaaaatctattgggagatgaagaaacacctctgtttggagacacttcaaaatagggc 4746  
Db 2087 gcatagaagaaatttggagtgaagataaaacctgttttggctcttacccttgcagac 2146  
QY 4747 agcataatgctcgaagacattaatgagtttgagttgcacagagagattacctgtccctc 4806  
Db 2147 tggatgactgtcttaggttcattgacagagatttgcgcagacacactggacagtggcatcag 2206  
QY 4807 aggtcctgtatcacagaanaacatctgtgttctcttatcagttgttcttcttaacataaaat 4866  
Db 2207 tttcagtggtgcaaggacatttttgaaccttttgcatacactggtgcctaataagacagcc 2266  
QY 4867 cagaagatacacacacatgccttctgtctatagatctgttttcaatgttttgggtggtctgtgt 4926  
Db 2267 atgagaaactccatgcataatagatatgacatgttttcatatttatttgggtgtgtgtg 2326  
QY 4927 tagcattcccatccttctatttgggagacactgtgtgctgcagcctcttctcagttagtt 4986  
Db 2327 ttgcatttctcgtgtgcagtg-----tcaggatttttcagggtatcagcc 2371  
QY 4987 ctctctataaccacctttatctcttccatttgcataccatggcacacatgcttcagatac 5046  
Db 2372 ttggcactggagacacttcacattttccatctgttactatgtgcacacatacagatct 2431  
QY 5047 taactacagtagaacac---aggcctaaccctgtcaggttccaagaagacagtgaagagg 5103  
Db 2432 tacttactcatgtacagaagagaatggcatggatcagaanaatcccccttgtgaagag 2491  
QY 5104 ctcatccgcacatctcttcttcttgcaaaaatttctcaatacacagtggtcccatgggt 5163  
Db 2492 aatcagcagttctgtgttcttataaaaacacttccaccagtaacaggaagtgccttgaag 2551  
QY 5164 gtgatatccctgctggtgatttgggtctcactgaagaatggcatcaccctctatcttc 5223  
Db 2552 aaataccatccgctggtcactgtgagagagtgacagagctggaatcatcgtcttctcctga 2611  
QY 5224 gctgtgctgcatgtgttttccactatttacttgggttaactccctccctgaggaactgcata 5283  
Db 2612 agtgttctgtttatttttttcttacttaaatggagtttctctcccccacccagatccaag 2671  
QY 5284 ccaattctgcagaagagagatagacagtcactctgtagctatctatcttaccatacaatt 5343  
Db 2672 t----tctggaaacaaagccatttgaacatttatgtagctatcttctccatcccaaacacc 2728  
QY 5344 tgttctcgtctctccaggaataatttgggtactgttaagggcccttgcctcagaggtggtgtg 5403  
Db 2729 tcaattgctcttttccaagaaaaatagtagataatgaattcactgatggaagtgtgtgccc 2788  
QY 5404 cagatcctgcttactaaactgtttgaagcaaaaaaacacccgtgttgaggtaccactaga 5463  
Db 2789 gtaacagtgagtttaaaagatatctagaagtgaaagagatgctataagatcatccaagag 2848





QY 3391 tagacacagatttcaacttgtggaccggaaggatgctccaaatggcgttttctatattctgg 3450  
Db 843 tggacaataatggtatctcgtggtcagagtcacgtcgcaagggtgttacatttaattg 902  
QY 3451 cattgggtttactagaagaagcaacagcttcacaaagctcctgaag---aagaagtaa 3507  
Db 903 gcatggcactacaagaagaaacacatttagagaatgtccggaagagcatgtagtaa 962  
QY 3508 catttgacttttataaggcttcaagattgggaagttcagccatgaat-----a 3558  
Db 963 catttaccttcaatcagaagatatcaaaacctggtgaagcgccaaaaattctctagca 1022  
QY 3559 tacaatgcttttggaaaaactocaaagaaattccccagttagaagccaggaagacatga 3618  
Db 1023 tactagctatgtggaacacactacaaaatgtccctacactagaagtcacacaagacatga 1082  
QY 3619 taactgatacttcagatgtttgacacagtgaaagcgatgaagcgttaagaacaaatctgttaa 3678  
Db 1083 ttcgggtgatatgaagacttttaatgctgttaaaagatgaggaga-----gttcac 1136  
QY 3679 ttgtagcaaccacatcaggaatggaattattaaagaatgagattactcatgataaag 3738  
Db 1137 ctaccagctcggtgcagagacagaaagaaacataatggaagagagttcaagggacaaag 1196  
QY 3739 aaaaagcagaacaaagaaagcgtgaagctgctaggtacatcgccagaaagacatgg 3798  
Db 1197 acaaagctlgagaggaagaaagcagagattgccagactgcccagagaaagacatgg 1256  
QY 3799 ctcaagatgctccttacagaaaaacttcattgaaactcataaaactcatgtatgacaata 3858  
Db 1257 ctcaagatgctgaaatgcagggcattttattgtgaaacaaagaaactcttccagcaga 1316  
QY 3859 catcagaatgctcgggaagaaagattccattatgaggaaagagagacaccagcagtcac 3918  
Db 1317 cattagaactggatgcctcaacctgtgttcttgatca-----tagccctgtggctt 1370  
QY 3919 gtgaactactagaattgctttgggtctctaaacgggttcactctgttactgaaaaggag 3978  
Db 1371 cagatgacacttacagacactggcccgccacaaactcaggttctctgaaacaaagacaat 1430  
QY 3979 tgcagctgcatctcttgcgaagaagaaacaggaaggtgaaatatagaaaaatagccatgg 4038  
Db 1431 tcgttacatgatatgtgtcaagaggagcaagaagttaaaatggaaagcagggcaatgg 1490  
QY 4039 tattatgcctgtgtccagaatctactgccttaccacagcacagggggaacacccatag 4098  
Db 1491 tcttggcagcatcttctcagagatcaactgtattatcaaaaaacagaagtaaatatttc 1550  
QY 4099 aactctcaggagaagcccttagaccacattttcatggatccagacttggcatatggaactt 4158  
Db 1551 aagatccagaaaaatat---gatccattattcatgacacctgactgtctgttggaaacac 1607  
QY 4159 atacagggaagctgtggtcatgtaagtacgcagctgtgtctggcagaagtatatttgaagctg 4218  
Db 1608 acactagtagctgtggcaccattatgcatgcccattgttggcaagggtatttggatccg 1667  
QY 4219 tacagctgaagctctcagcagcgcattcatgttgacactttt-----tgacttgg 4266  
Db 1668 tccaagctaaagaacaggaaggaacagagattacgcttatcatagcagctatgtagtag 1727  
QY 4267 aaagtggagaaatatcttgcctctcttgcacaaactctgtgcataactgtgatcccatca 4326  
Db 1728 aaaaaggagaattctcttgcctcttgcctcttgcctcttgcctcttgcctcttgcctctg 1787  
QY 4327 ttcctttgcaacctcaaaagataaacagtgagaatgcagatgctcttgcctcaacttttga 4386  
Db 1788 tgcctt-----ctcaagaaatatattttaacacaggttaaaatttttcagaccaaccaa 1841  
QY 4387 ccttggcagcgttggtatagacagactgttctggccagaaatatacaggttataataagaacatg 4446  
Db 1842 atctgactcagtggttagaacaatatctcagcaaaataaaagcattacattgttcttagga 1901  
QY 4447 ctaaaggagaaaaaccaaattcctatttctttaaatacaaggaaatggagattctacttgg 4506

Db 1902 aagaagaaagtactccttaataatgctctctacaaagaattcagaaaaatgtggatgaattac 1961  
QY 4507 agtctattccatcctcagatttttggcgttgagctcttcagataataatctcaaatagcatca 4566  
Db 1962 agtccctgaagggttcaggcctgattttcgtcctaagatccccctattcttgagagataa 2021  
QY 4567 aggaataatgttattctcttgcacacaataattatagaattggattgaaagtgcacactg 4626  
Db 2022 aagaataatgctaaagacatttgggaactgctaccacaaaggtgggaactaaagttcatcca 2081  
QY 4627 atgaaggagatcctcagatgctcccatgctgaccttggagacacactgcgtcttcaactaccag 4686  
Db 2082 atgaaggagatcctcgtgttcccatatgttgggtagctgcgtcgcgtacacatccaaa 2141  
QY 4687 caattgaataatctatigggagatgaagaaacactctgttggagcacttcaaaaataggc 4746  
Db 2142 gcatagaagaatttggagtgatgaagataaaacattgttggctccttacccttgcagac 2201  
QY 4747 agcataatggtctgaagcatttaagtgcagttgcagttgcacagagagattacctgtctc 4806  
Db 2202 tggatgactgtcttaggtcattgacgagatttgcgcgacacactggacagtggtcatcag 2261  
QY 4807 aggtcctgatcacagaacacactcgttgcgttctctatcagattgttcttccctaaacataaaat 4866  
Db 2262 tttcagtggtgcgaagacatttttgaaccttttgcacactgttgcctaaatgacagcc 2321  
QY 4867 cagaagatacacacactcctctctctatagatctgttcttcttcttgggtgggtgctgtgt 4926  
Db 2322 atgaggaaactccatgcataattagatattgacatgttcttatttattgttgggtgggtgc 2381  
QY 4927 tagcattcccatcctgtatttggatgacctgttgatctgcagcctcttctcagttagtt 4986  
Db 2382 ttgcatttctcgttgcagtg-----tcaggatttttcagggacagcc 2426  
QY 4987 ctctctataaccaccttattctcttccatttgcacacttgcagagacacactgcttcagatca 5046  
Db 2427 ttggcactggagaccttcacatttccatctgttactatggcacacatcatacagatct 2486  
QY 5047 taottacagtagacac---aggcctacctctgctcaggttccaaagaagacagtgaaagg 5103  
Db 2487 tactacctcatgtacagaagaagatggcatgatacagaataatccccccttggagaag 2546  
QY 5104 ctctaccgcatctcttcttgcagaaatttctcaatatcaaatgtgctccacttgggt 5163  
Db 2547 aatcagcagttctgttgtatataaaacacttccacagtatcacgggaagtccttgaag 2606  
QY 5164 gtgatattcctggtggtatttgggtctcactgaagaatggcatcaccccttatcttcc 5223  
Db 2607 aaataccatccggtgcatctgtggaggaggttcagagctgggaatcatgcttctctga 2666  
QY 5224 gctgtgctgcatgttttccactatttacttgggttaactccgctcaggaagacatgata 5283  
Db 2667 agtctcgtcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttct 2726  
QY 5284 ccaattctcagaagagagatcacagtcactctgtagctatcttcttcttcttcttcttct 5343  
Db 2727 t---tctggaaacagccattttgaaacattttagtagctatcttcttcttcttcttcttct 2783  
QY 5344 tgttctgctcttccagggaataatttgggatactgttaagcccttgcctccagaggtgtgtgt 5403  
Db 2784 tcaattgcttcttcaagaaataagtgagataatgaattcactgattgaaagtgtgtgtgt 2843  
QY 5404 cagatcctgcttactaaactgtttgaagcaaaaaacacccgttggtcaggttaccctagaa 5463  
Db 2844 gtaacagtggaagttaaagatatctagaagtgaagagatgctataagatatccaagag 2903  
QY 5464 aagaataatgtttagagagcttccctgatactatagctgctcctgaatacaagctctc 5523  
Db 2904 aatcaacaaataaataacccctccagagagattacagcagcctcatataatcaagcatcca 2963  
QY 5524 atttcaggtgcccacggctgtgcagatgatgagcgaaagatccctgtcctctgctcttct 5583

Db 2964 atttctgtgccgaatacagggtggtataagagacagagcccaactctgtgccttggt 3023  
QY 5584 gtggggtactactgttctcagacaatttctgtccaggaataattgtgacgggggaagagg 5643  
Db 3024 gcggtctctgtgtctccagagttactgtctgcagactgaactggaaggaggagtg 3083  
QY 5644 ttggagcttcatttttcagccacttcactgtgagcggagctgctgaattttcctaaaa 5703  
Db 3084 taggagctgcacagccacactactctctgtgctctggtgagtggtggtcctctcagag 3143  
QY 5704 tcagagaatgcgagtggtctgtggtgaaggtaaagccagaggtgtgctcctatccagctc 5763  
Db 3144 taggggaatgcagtgctatttttagctggaacaaacaaaggctgttttattctctc 3203  
QY 5764 cttacttgatgaatatgggaaacacagaccctggtcctggaagggggaacccctcatt 5823  
Db 3204 cttactctgactatgggagaccagggactcagcggggaatcctttacatt 3263  
QY 5824 tatctctgagcgtatcggaagctcatttggctgtggtcgaacaaacactgcattataga 5883  
Db 3264 tatgcaaaagcgtattcaagaagattcagaagctctggcaccacacacagtgctcacag 3323  
QY 5884 agattgctaggagccaagagactaatcagatgttatttggattcaactggca 5935  
Db 3324 aaattgacatgcacaggaagccaatcagacactggttggcattgactggca 3375  
RESULT 8  
AAA02411  
ID AAA02411 standard; cDNA; 733 BP.  
XX  
AC AAA02411;  
XX  
DT 19-MAY-2000 (first entry)  
XX  
DE Human colon cancer cell line polynucleotide sequence SEQ ID NO:2402.  
XX  
KW Human; colon cancer; tumour; diagnosis; gene expression product;  
KW probe; detection; cancerous state; metastasis; identification;  
KW breast cancer; oestrogen receptor-positive breast cancer; therapy;  
KW oestrogen receptor-negative breast cancer; lung cancer; ss.  
XX  
OS Homo sapiens.  
XX  
XX WO9558675-A2.  
XX  
XX PD 18-NOV-1999.  
XX  
XX PF 13-MAY-1999; 99WO-US10602.  
XX  
XX PR 14-MAY-1998; 98US-0085426.  
XX PR 15-MAY-1998; 98US-0085537.  
PR 15-MAY-1998; 98US-0085696.  
PR 21-OCT-1998; 98US-0105234.  
PR 27-OCT-1998; 98US-0105877.  
XX  
XX (CHIR ) CHIRON CORP.  
PA (HYSE-) HYSEQ INC.  
XX  
XX Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;  
PI Reinhard C, Giese K, Randazzo F, Kennedy GC, Pot D, Kassam A;  
PI Lamson G, Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I;  
PI Leshkowitz D, Kita D, Garcia V, Jones IW, Stache-Crain B;  
XX  
XX WPI; 2000-126369/11.  
XX  
XX Polynucleotide library used to determine cancerous states of mammalian  
PT cells -  
PT  
XX  
XX Claim 1; Page 958; 1097pp; English.  
PS  
XX  
CC AAA00010 to AAA02716 represent polynucleotides isolated from cDNA  
CC libraries constructed from human colon cancer cell lines. The present

CC invention also describes a method of detecting differentially expressed  
CC genes correlated with a cancerous state of a mammalian cell, comprising  
CC detecting at least one differentially expressed gene product in a test  
CC sample derived from a cell suspected of being cancerous, where detection  
CC of the differentially expressed gene product is correlated with a  
CC cancerous state of the cell from which the test sample was derived.  
CC The polynucleotides sequences can be used in a method for detecting  
CC differentially expressed genes correlated with a cancerous state of a  
CC mammalian cell. The polynucleotides can also be used as probes for  
CC detecting and mapping related genes. They can be used in diagnosis and  
CC prognosis of diseases and disorders (e.g. identification of  
CC pre-metastatic or metastatic cancerous states, stages of cancer, or  
CC responsiveness of cancer to therapy). This is particularly for breast  
CC cancer, oestrogen receptor-positive breast cancer, oestrogen receptor-  
CC negative breast cancer, lung cancer, and colon cancer.  
XX  
SQ Sequence 733 BP; 218 A; 145 C; 152 G; 210 T; 8 other;  
  
Query Match 10.4%; Score 656.8; DB 21; Length 733;  
Best Local Similarity 97.5%; Pred. No. 7e-169;  
Matches 696; Conservative 0; Mismatches 13; Indels 5; Gaps 3;  
  
QY 1405 catatgaccacgtcatatatacagcctacaagagctcttgactgtgagtcgcagagccc 1464  
Db 20 cagcagaccacgtcatatatacagcctacaagagctcttgactgtgagtcgcagagccc 79  
QY 1465 agttgcataccactgcattgacaaagagggtcgctgggctgttaaagcgaggacttatg 1524  
Db 80 agttgcnttccactgcatgacaaagagggtcgctgggctgttaaagcgaggacttatg 139  
QY 1525 ctgcttgcaggagcaaggaagatatataagagtgcttcacagaaaatgtctctcaacatc 1584  
Db 140 ctgcttgcaggagcaaggaagattataagagtgcttcacagaaaatgtctctcaacatc 199  
QY 1585 cacttcagttagaagtattacactcagagattatggctcactcagaaaatgtctctcagc 1644  
Db 200 cacttcagttagaagtattacactcagagattatggctcactcagaaaatgtctctcagc 259  
QY 1645 ttggtctctggatgaacaaattatagctattcaagtgactttaggcagatcttttgcc 1704  
Db 260 ttgg-tcttgatgaacaaattatagctattcaagtgactttaggcagatcttttgcc 318  
QY 1705 aagcatgcttagaagaacactgactcggagataccctgtctcactaagcagggttaatgc 1764  
Db 319 aagcatgcttagaagaacactgactcggagataccctgtctcactaagcagggttaatgc 378  
QY 1765 ttgggagtcgaagctttataaagggtgcgcgtaagatccttcctatgaattgacttcagca 1824  
Db 379 ttgggagtcgaagctttataaagggtgcgcgtaagatccttcctatgaattgacttcagca 438  
QY 1825 gttttttatgagatggaatacaaaaaactctttgtctatggaattgtgaagtattata 1884  
Db 439 gttttttatgagatggaatacaaaaaactctttgtctatggaattgtgaagtattata 498  
QY 1885 aacaactgcagaagaataatatacagtgtatgatcatgacagaagtatctctataactcac 1944  
Db 499 aacaactgcagaagaataatatacagtgtatgatcatgacagaagtatctctataactcac 558  
QY 1945 ttctcagttcagatgtttactgttctactctctgtgctgcacatcttattgaagagcagaatg 2004  
Db 559 ttctcagttcagatgtttactgttctactctctgtgctgcacatcttattgaagagcagaatg 618  
QY 2005 ttatctctgctactgaactcgtcagaagttttacotcgtactgtgagcagg--aa 2062  
Db 619 ttatctctgctactgaactcgtcagaagttttacotcgtactgtgagcagg--aa 678  
QY 2063 caataaaattcaacttccagggttatagcc--aggacaaaattgggaagagtatat 2114  
Db 679 caataaaattcaacttccagggttatagcc--aggacaaaattgggaagagtatat 732  
  
RESULT 9







Db	639	TTAAAAGATTCTTAGAAGGTGAAAGAGATGCTATTAGATATCAAGAGAAATCTTACAA	580
QY	5474	tttgatagagcttctctgatgactatagctcctcctcctgaatcaagcttctcatttccaggtg	5533
Db	579	ATTATAAACCCTCCAGAGGATTACAGCAGCCTCATTAATCAAGCATCCAATTTCTCGTG	520
QY	5534	cccacggtctcagatgatgagcgaaagcatcctgtcctcgtcctcgtccttctgtggggctat	5593
Db	519	CCCGAAATCAGGTGGTGTGATAAGAGCAGAGCCCAACTCTGTGCCTTGTGCGGATCTCT	460
QY	5594	actatgtcttcagaacatttctgcgcagagaaattgtgaacgggaaagattgagagcttg	5653
Db	459	GCTGTGCTCCCAAGTTACTGCTGCCAGACTGAACCTGGAAGGGAGGATGTAGGAGCCTG	400
QY	5654	cattttccacgacttcaactgtgagccgagctctgcattttcctaaaaatcagagaatg	5713
Db	399	CACAGCTCACACCTACTCTCTGTGGCTCTGGAGTGGGCATCTTCTGAGAGTACGGGAATG	340
QY	5714	ccgagtggtcctcgttgaaaggttaaacgagaggtgtgcctatccagctccttacttggga	5773
Db	339	TCAGGTGCTATTATTTAGCTGGCAAAACCAAGAGCTGTTTTATTCTCTCTCTACCTTGA	280
QY	5774	tgaatatgagaaacagaccctggtcctgaagaggggcaaccccttcatttactctctga	5833
Db	279	TGACATATGGGGAGACCGACAGGACTCAGACGGGAAATCCTTTACATTTATGCAAGA	220
QY	5834	gcggtatcggaagctccatttggctcgtggaacaacacactgcattatagaagagattgctag	5893
Db	219	GGATTCAGAAGATTCAGAAGCTCTGGCACCAACACAGTGTACACAGAGAAATTTGGACA	160
QY	5894	gagccaagagacatacagatgttatgttgattccaactggca	5935
Db	159	TGCACAGGAAGCAATCAGACACTGGTTGGCATTTGACTGGCA	118
RESULT	14		
AAV89674/c			
ID	AAV89674	standard; cdNA; 455 BP.	
XX			
AC	AAV89674;		
XX			
DT	15-FEB-1999	(first entry)	
DE			
OS	EST clone DA490.		
XX			
KW	Human; secreted protein; expressed sequence tag; EST; haematopoiesis;		
KW	tissue growth; activin; inhibitor; chemotaxis; chemokinesis; haemostatic		
KW	receptor; ligand; thrombolytic; anti-inflammatory; cadherin; anti-tumo		
XX	gene therapy; ss.		
OS	homo sapiens.		
XX			
PN	W09845436-A2.		
XX			
PD	15-OCT-1998.		
XX			
PF	10-APR-1998; 98WO-US06955.		
XX			
PR	10-APR-1997; 97US-0838821.		
XX			
PA	(GEMV ) GENETICS INST INC.		
XX			
PI	Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;		
PI	Racie LA, Spaulding V, Treacy M;		
XX			
DR	WPI; 1999-070077/06.		
XX			
PT	New polynucleotides encoding human secreted proteins - derived from		
PT	e.g. human blood, kidney, foetal lung, placenta, testes, brain,		
PT	ovary, pituitary, retina and colon cdNA libraries.		
XX			
PS	Claim 1; Page 293; 618pp; English.		
XX			

CC The present sequence represents a human expressed sequence tag (EST).  
 CC The polynucleotide, which is a secreted EST, and the encoded protein  
 CC are predicted to have useful biological activities which would make  
 CC them suitable for treating, preventing or ameliorating medical  
 CC conditions in humans and animals, although no supporting data is  
 CC given. Suggested activities include nutritional activity, immune  
 CC stimulating or suppressing activity, haematopoiesis regulating  
 CC activity, tissue growth activity, activin/inhibin activity,  
 CC chemotactic/chemokinetic activity, haemostatic and thrombolytic  
 CC activity, receptor/ligand activity, anti-inflammatory activity,  
 CC cadherin/tumour invasion suppressor activity, tumour inhibition  
 CC activity. The polynucleotide may also be useful for gene therapy.  
 XX  
 SQ Sequence 455 BP; 108 A; 121 C; 107 G; 119 T; 0 other;

Query Match	2.5%	Score 159;	DB 20;	Length 455;
Best Local Similarity	61.9%;	Pred. No. 4.3e-33;		
Matches 252;	Conservative 0;	Mismatches 155;	Indels 0;	Gaps 0;
QY 5529	agggtccacggtctgcagatgatgagcgaaagcatcctgtccctctgcctttctgtggg	5588		
Db				
QY 451	AGGTGCCGGAATCAGGTGGTATTAAGAGCAGAGCCCACTCTGTGCTTGTGTGCGGA	392		
Db				
QY 5589	gctatactatgtttctcagaacatttgctgcagagaaattgtgaacgagggaagaggttggga	5648		
Db				
QY 391	TCTCTGTGTGCTCCACAGGTTACTGTGTGCAGACTGAACTGGAAGGGGAGGATGTAGGA	332		
Db				
QY 5649	gcttgcaattttcaagcacttcactgtgtgagcggaggtctgcattttctctaaaaatcaga	5708		
Db				
QY 331	GCCGTGCACAGCTACACCTACTCTGCTGGCTCTGGAGTGGGCATCTTCTCTGAGAGTACGG	272		
Db				
QY 5709	gaatgcgcagtggttcctcgttggtgaaggtaaagccagaggtcgtgcctatccagctcctcac	5768		
Db				
QY 271	GAATGTCAGGTGCTATTTTATAGCTGGCAAAACCAAGGCTGTTTATTCTCCTCCTTAC	212		
Db				
QY 5769	tggatgaatatgagaaaaacagacctggcctgaagaggggcaaccccccttcattatct	5828		
Db				
QY 211	CTTGATGATTATGGGAGACCGCAGCGGACTCAGACGGGAAATCCCTTTACATTTATGC	152		
Db				
QY 5829	cgtgacggtatcggaagctccatttggctcggcaacaaactgcattatagaagagatt	5888		
Db				
QY 151	AAAGAGCGATTCAAGAGATTCAGAAAGCTCTGGCCACCACACAGTGTCCAGAGGAAATT	92		
Db				
QY 5889	gctaggagcaagagactaatcagatgttatgttgattcaactggca	5935		
Db				
QY 91	GGACATGCACAGGAAGCAATCAGACACTGGTGGCATTTGACTGGCA	45		
Db				

RESULT	15	
AAF58252/c		
ID	AAF58252	standard; DNA; 936 BP.
XX		
XX	AAF58252;	
XX		
DT	24-APR-2001	(first entry)
XX		
DE	Oligonucleotide	D1835.
XX		
KW	Electron-transfer group;	ETM; mismatch; genotyping;
KW	gene expression;	ss.
XX		
XX	Synthetic.	
XX		
PN	WO200107665-A2.	
XX		
PD	01-FEB-2001.	
XX		
PF	26-JUL-2000;	2000WO-US20476.
XX		
XX	26-JUL-1999;	99US-0145695.
PR	17-MAR-2000;	2000US-0190259.
PR		
XX		

PA (CLIN-) CLINICAL MICRO SENSORS INC.  
XX  
XX Umeek RM;  
XX  
XX  
XX WPI; 2001-159728/16.  
XX  
XX Nucleic acids containing electron-transfer group, useful as labels in  
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on  
PT a single surface  
XX  
XX  
XX Example 6; Page 127; 159pp; English.  
PS  
XX  
XX The present invention relates to a composition comprising two nucleic  
CC acids each containing an electron-transfer group (ETW) having  
CC different redox potentials. The invention is used for electronic  
CC detection of nucleic acids, especially of substitutions (mismatches)  
CC and single-nucleotide polymorphisms, e.g. for genotyping,  
CC monitoring gene expression.  
XX  
XX Sequence 936 BP; 4 A; 139 C; 10 G; 7 T; 776 other;

[illegible]





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 1, 2002, 12:40:50 ; Search time 95.27 Seconds  
(without alignments)  
14995.520 Million cell updates/sec

Title: US-09-724-126A-1  
Perfect score: 6308  
Sequence: 1 gccagaattcgccagcagg.....aatttgatttggtgtttt 6308

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA:\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PTUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4435.4	70.3	6395	2	US-08-982-956-1
2	4435.4	70.3	6395	3	US-09-228-317-1
3	996.2	15.8	1001	2	US-08-982-956-2
4	996.2	15.8	1001	3	US-09-228-317-2
5	57.4	0.9	7218	1	US-08-232-463-14
6	52.2	0.8	7218	1	US-08-232-463-14
7	44.4	0.7	7333	3	US-09-138-024-21
8	44.2	0.7	72928	3	US-09-009-913-1
9	42.2	0.7	325	4	US-08-991-789A-236
10	41	0.6	1517	3	US-08-963-602-6
11	40.4	0.6	152331	3	US-09-128-155-16
12	39.8	0.6	2861	1	US-08-299-953-1
13	39.8	0.6	2861	1	US-08-459-415-1
14	39.8	0.6	2861	5	PCT-US95-11231-1
15	39.8	0.6	3881	1	US-08-299-953-2
16	39.8	0.6	3881	1	US-08-459-415-2
17	39.8	0.6	3881	5	PCT-US95-11231-2
18	39.8	0.6	4376	1	US-08-119-125A-1
19	38.8	0.6	246240	2	US-08-724-394A-20
20	38.8	0.6	246240	2	US-08-724-394A-21
21	38.8	0.6	246240	2	US-08-724-394A-22
22	38.4	0.6	5703	4	US-09-280-590A-36
23	38.4	0.6	18596	4	US-09-318-448-11
24	37.8	0.6	19011	1	US-08-310-356-36
25	37.8	0.6	19557	5	PCT-US92-06300-1
26	37.6	0.6	72928	3	US-09-009-913-1
27	37.4	0.6	1333	4	US-08-543-246B-15

c 28	37.4	0.6	1387	4	US-08-543-246B-1	Sequence 1, Appli
c 29	37	0.6	1113	1	US-08-341-538A-1	Sequence 1, Appli
c 30	37	0.6	1113	2	US-08-725-518-1	Sequence 1, Appli
c 31	37	0.6	1618	3	US-08-889-108-1	Sequence 1, Appli
c 32	37	0.6	1618	3	US-08-889-108-3	Sequence 3, Appli
c 33	37	0.6	1618	4	US-08-120-601B-1	Sequence 1, Appli
c 34	37	0.6	1618	4	US-08-120-601B-3	Sequence 3, Appli
c 35	37	0.6	1618	5	PCT-US94-10358-1	Sequence 1, Appli
c 36	37	0.6	1618	5	PCT-US94-10358-3	Sequence 3, Appli
c 37	37	0.6	4875	1	US-08-460-739-1	Sequence 1, Appli
c 38	36.6	0.6	1480	4	US-09-290-640-65	Sequence 65, Appli
c 39	36.6	0.6	3784	1	US-07-623-033-1	Sequence 1, Appli
c 40	36.4	0.6	1222	4	US-08-543-246B-5	Sequence 5, Appli
c 41	36.2	0.6	9636	1	US-08-323-170B-1	Sequence 1, Appli
c 42	36.2	0.6	19124	2	US-08-487-826B-13	Sequence 13, Appli
c 43	35.4	0.6	252	4	US-08-332-766A-28	Sequence 28, Appli
c 44	35.4	0.6	2135	4	US-08-430-286A-1	Sequence 1, Appli
c 45	35.2	0.6	289	4	US-09-007-005-17	Sequence 17, Appli

ALIGNMENTS

RESULT 1  
US-08-982-956-1  
; Sequence 1, Application US/08982956  
; Patent No. 5861312  
; GENERAL INFORMATION:  
; APPLICANT: Varshavsky, Alexander  
; APPLICANT: Kwon, Yong Tae  
; TITLE OF INVENTION: NUCLEIC ACID ENCODING MAMMALIAN UBRI  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Kevin M. Farrell, P. C.  
; STREET: P.O. Box 999  
; CITY: York Harbor  
; STATE: ME  
; COUNTRY: US  
; ZIP: 03911  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/982,956  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Farrell, Kevin M.  
; REGISTRATION NUMBER: 35,505  
; REFERENCE/DOCKET NUMBER: CIT-2001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (207) 363-0558  
; TELEFAX: (207) 363-0528  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 6395 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: CDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 115..5385  
; US-08-982-956-1

Query Match 70.3%; Score 4435.4; DB 2; Length 6395;  
Best Local Similarity 86.3%; Pred. No. 0;  
Matches 4997; Conservative 0; Mismatches 721; Indels 69; Gaps 6;  
Oy 590 ttcaggggccgctgtaaaagtgtcgtccctgtctc-tccgaccggccacaggtttccgct 648



Qy	2809	ataagttcttgtaactgttaacttcagaggatagaacttgcgagggctttttaacaaagacca	2868
Db	2228	ACAAGTCTTGTTACTGTGTACTTCAGAGATATCAACTTACTGATGCTTTTAACAAGACCA	2287
Qy	2869	tatctcaaaaagaccaggaattgattaaacaataataacactaatagaagaatacttcttc	2928
Db	2288	TATCCACAAAAGACCAGGATTTGATTAAACACTATATACATTAAATAGAAGAAATGCTTC	2347
Qy	2929	aggctctcatctatatgttggtggagcgttatgtacctggagtgggaaatgtgaccaaag	2988
Db	2348	AGGTCTCATCTATATTGTGGAGAACGTTATGTACTCGAGTGGGAAATGTTTACCAGAG	2407
Qy	2989	aagaggtcacaaatgagaaaatcattcacttcttggattgaacccatgccacacagt	3048
Db	2408	AGGAGTTATATGAGAGAGATTACTACTTCTTGTGCATTGAGCCCATGCCACACAGTG	2467
Qy	3049	ccattgccccaaatttacctgagaatgaaaaaataagaactggcttagagaaatgtcataa	3108
Db	2468	CCATCGCCAGAAACCTACTCTGAGACGAAAAATTAATCAAACTGGCTTAGAGAAATGTCATA	2527
Qy	3109	acaaagtgccccacatttaagaacacagggtgatacaggccatggagtttatgaactaaaag	3168
Db	2528	ACAAAGTGGCCACATTTAAGAAACACAGGTGTCTCGGCCCATGGAGTTATGAATGAAAG	2587
Qy	3169	atgaatcaactaaaagacttcaatatgtactttttatcatttaccataaaacccagcatagca	3228
Db	2588	ATGAATCACTGAAGACCTTCAATATGTACTTTTACATTATTCTAAACACACAGCATAGCA	2647
Qy	3229	aggctgaacatatgcagagaagaaggaaacaagaacaaagaatgaagcattgccgc	3288
Db	2648	AGGCTGAACATATGCAGAAGAAAAGGAGAAAACAGAAATAAAGATGAAGCATTCGCCG	2707
Qy	3289	caccacacctctgaaattctgcctgtttcagcaaatgtattaaccttctcaactgtg	3348
Db	2708	CGCCACCTCTCCAGAGTTCTGCCCTGTCTTCAGGAAAGTAGTCAACCTGCTCAGCTGTG	2767
Qy	3349	atatcatgatatacttcaggaacggtattgagcgggcaatagacacagattctaaact	3408
Db	2768	ATGTTATGATATACATCTCTAGGACCATCTTTTGGCGGCAGTGGACACGGAGTCTAATC	2827
Qy	3409	tgtgaccgaagggtgcctccaaatggcttttcattcttgcattgggttttactagaag	3468
Db	2828	TGTGCACAGAAGGGATGTGCAGATGGCGTTCATATATTGCACTGGCGTTGCTGGAAG	2887
Qy	3469	agaagcaacagcttcaaaaagctcctgaagaagaagtaacaattgaactttatcataag	3528
Db	2888	AGAAGCAGCAGCTTCAGAAAGCTCTGAAGAGGAAGTGCTTTTGACTTTTACCATTAAG	2947
Qy	3529	cttcaagattgggaagttcagccat-----gaatatcaaatgcttttggaaaaac	3579
Db	2948	CTTCAAGATTGGGAAGTTGAGCCATGAATGCTCAGAAATATACAATATGCTCTTGGAAAGAC	3007
Qy	3580	tcaaggaattccccagttagaagccagaaggaatgataacgttggaacttccagatgt	3639
Db	3008	TCAAAGGAATCCCCAATTAGAAGGCCAGAGACATGATAACATGGATACTCCAGATGT	3067
Qy	3640	ttgacacagtgaagcgaataagagaaaaatcttgtttaatgtgacaacacacatcaggat	3699
Db	3068	TTGACACAGTGAAGCGGATTAAGAGAAAAATCTTGTTTACTTGTGGCAACCACTTCCAGGAC	3127
Qy	3700	cggaatctttaagaatgatagattactcatgataaagaaaaagcagaacgaagaaagaa	3759
Db	3128	TGGAGTGATTAAAGAGTGAAGGAGATTACTTCATGATTAAGAAAGGCAAGCAAGCAAGAA	3187
Qy	3760	aagctgaagctgctaggctacatgccagaagaatcatggctcagatgtctgcctttacaga	3819
Db	3188	AAGCTGAGCCGCTAGGCTTCATCGCCAGAAACATCATGCCCCCAGATGCTGCTTTACAGA	3247
Qy	3820	aaaacttcaactcataaactcatgtatgacaatatcatagaataatgccttgggaaag	3879
Db	3248	AAAACTTCATTGAACCCACAAACTCATGTATGTATGATTAATACGTACAGAAGTAACAGGGAAG	3307

Qy	3880	aagattccaattatgaggaagagagacacccagcagctgactactctagaaattgcttt	3939
Db	3308	AAGACTTCCATTATGAGGAAAGAGAGACCTCAGCAGCTCAGTGTAGAGGCTCTAGAAATTGCTC	3367
Qy	3940	tgggtcctaacaacggggtccattctgtacttgaataagagggtgtctgacgctgcatcctttggc	3999
Db	3368	TGGGCCCTTAACGGGGCCGGCTGTATACCGAAAAGAGGTTGCTTACGTGCGATCCTCTGCC	3427
Qy	4000	aagaagaacagggttgaaaaatagaaaaataatgccccattgtattatcggccgtgtgtccaga	4059
Db	3428	AAGAAGAACAAGAGGTAAACTAGAAAAATAATGCCATGGTATTGTTCAGCATGTGTGCAGA	3487
Qy	4060	aattctactgctttaaccacagacagagggaaaaaccatagaactctcaggagaagccctag	4119
Db	3488	AATCACCAGCCCTAACCCAGCAGAGAGGAAGGCTGTGACCACCTATTAGGGGAACACTGG	3547
Qy	4120	accacttttcatggtatccagacttggcattatggaacttatacagaagctgttgggtcatg	4179
Db	3548	ACCCCTTTTTCATGATCCAGACTTGGCACAATGGAACCTTATACAGGAAGCTGTGGTCATG	3607
Qy	4180	taatgcacgcagctgtctggcagaaagtattttgaagctgtacagctgagctctcagcagc	4239
Db	3608	TAATGCATGCAGTGTCTGGCAGAAGTATTTTGAAGCTGTGACAGCTGAGCTCGCAGCAGC	3667
Qy	4240	gcattcatgtgacotttttgacttggaaagtggaataatatctttgccctctttgcaaat	4299
Db	3668	GCATTCACGTAGACCTGTGTTGACTGGAGAGCGCGAGTACCTTATGTCGCCCTCTGCAAGT	3727
Qy	4300	ctctgtgaataactgatcccccattatcccttggaaactccttgaacctcaaaagataaacagtgaga	4359
Db	3728	CTCTCTGC AACACTGTATCCCATCATCCCTTTGACGCCGAGAGATCAACAGTGAAGA	3787
Qy	4360	atgcagatgctcttgcctaaccttttgacccctggcagcggtggatcacagactgttcttggcca	4419
Db	3788	ATGCGGAGGCTCTTGCTCAACTTTTGACCTTGGCAGCGCGAGTACCTTATGTCGCCCTCTGCAAGT	3847
Qy	4420	gaatcatagggttataataataagacatgctaaaggagaaaaacc---caattccctattttct	4476
Db	3848	GAATATCGGGTTTATATATPAAAGCATGCTPAAAGGAGAAGCCCCCAGCAGTGTCCCTGTCTTGT	3907
Qy	4477	ttaatcaaggaaatggagattctactttgagttccattccactccctgagtttttggcgttg	4536
Db	3908	TTAATCAGGAATGGGGATTCAACTTTTGAGTTTCATTCCTCCATCCGTGAGTTTGGAGTTC	3967
Qy	4537	agctctcgattaaatattcaaatagcatcaaggaataatggttattctcttttggcacacaacaa	4596
Db	3968	AGTCITCGGTGAATATCAAAATAGTATCAAGGAATGGTCAITCTCTTCGCCACAACAA	4027
Qy	4597	tttatagaatttgatgaaagtgcacactgatgaaaggatcctcagtgccccatgctga	4556
Db	4028	TTTACAGAAATTGGCCTGAAAGTGCCTCTCTGATGAACCTAGACCCACGAGTGCCCATATGA	4087
Qy	4657	cctggagcacctgcgctttcactatccagcaattgaaaaatctattggagatgaaggaa	4716
Db	4088	CTTGAGCACGTGTGCGTTTCCATCTCCAGGCAATCGAAACCTGTGTGGGAGATGAAGGAA	4147
Qy	4717	aacctctgtttgagcaccttcaaaataggcagcataatggctgtgaaagcattaatgcagt	4776
Db	4148	AACTCTATTTGGACACTTCAAAATAGACGATAGCGGTCTGAAGGCCCTTAATGCAGT	4207
Qy	4777	tgcagttgcacagaggattacctgtccctcaggtcctgatcacagaacaatctcgttctgtc	4836
Db	4208	TTGACGTTGCACAGAGGGCTACCTGCGCCTCAGGTCTGTATACACAAACACTCTGGCTCGGC	4267
Qy	4837	ttctatagttgttcttctcaataaaaaatcagaagatatcacacatgccttctgtctatag	4896
Db	4268	TCTCTCAGTTATCTTCTTAACCTGCAATTCAGAAATAACACAGGCCCTTCTCTCTGTGG	4327
Qy	4897	atctgtttcatgttttgggtgctgtgttagcatcccatccttgtattggatgacc	4956
Db	4328	ATCTCTTCATCTTCTGTGTCGGCGCAGTCTTACGCTTCCATTCCTTGTATTGGATGACA	4387
Qy	4957	ctgtttgatctgcagcctctcttcagtttagttcttctcctaacaaccctttatctctccatt	5016

Db 4388 CCGTGGATCTGCAGCGCTCGCACCTTAGTCTCTCATATAAACACCTCTATCTCTTCCATC 4447  
Qy 5017 tgataccatggcacacatgcttcagatactactactactacagtagaca-----cag 5064  
Db 4448 TGATCACCATTGGCGCACATGCTTTCAGATACCTCTTACACACAGATACAGATCTGTCTCCAG 4507  
Qy 5065 gacctacccttgctcaggttcaagaagacagtgaaagagcctcaatccgcattcttctct 5124  
Db 4508 GCGCCCGCTGCTGAGGGTGAAGAGGATAGTGAAGAGCGCTGCTGTGCATCTGCTTCT 4567  
Qy 5125 ttgcagaatttctcaalatacaagtggctccatgggtgtgatatctctcgtgctgatt 5184  
Db 4568 TTGTGGAGTGTGCGACGACACAGACGCGCTCAGTGGGTGCGGTCTCCGCGCTGGTACC 4627  
Qy 5185 tgtgggtcactgaagaatggcacaaccccttatctctgctgtgctgattgttttcc 5244  
Db 4628 TGTGGCTCTCCGTAGGAAGCGCATCCCGCTTACCTCCGCTGTGCTGCACTGCTTTTCC 4687  
Qy 5245 actatttacttgggttaactccgctgaggaactgcataccaattctgcagaaggagat 5304  
Db 4688 ACTATTACTTGGAGTAGTCCGCGTGAAGAACTGTTTGCCAAATCTGCTGAAGGAGAT 4747  
Qy 5305 acagtgcactctgagctatctattacatcaaaattgttctcctgctctccaggaat 5364  
Db 4748 TCAGTGCACCTGTGAGCTATCTATCTTTACCCACAAAATTGTTCTGCTTTTCCAGGAAT 4807  
Qy 5365 attgggtactgaaggcccttgcctccagagtggtgagatcctgcttactactaaact 5424  
Db 4808 ATTGGATACCAATAAGGCCCTTACTACAGAGGTGCTGGAGACTCTGCTTACTCAAGT 4867  
Qy 5425 gtttgaagcaaaaaaacccgtggtgcaagtaacctagaaaaaagaataagttgatagac 5484  
Db 4868 CTTTGAAGCAGAAAAAGTGTGTGGTACAGGTACCTAGAAAAAGAAATAGTTTGATAGAC 4927  
Qy 5485 ttcctgtagactatagctgctccctgaatcaagcttctcattcagggtgccacaggtcg 5544  
Db 4928 TTCTTGAGACTACAGCTGTCTTAAATCAGGCTTCTCACTTAGGTGTCCACGGCTCG 4987  
Qy 5545 cagatgtagcgaagacatcctgctcctctgctcttcttctggtgggctatactatgtctc 5604  
Db 4988 CAGATGATGAGGCAAGCATCTCTCTCTGCTTCTTCTGTTGGGCCATCTGCTGTCTC 5047  
Qy 5605 agaacttgcctgcaggaataattgtaacggggaagaggttgagagctgctattttcag 5664  
Db 5048 AGAACATCTGTGCGCAAGAAATAGTGAATGGGAAGAGGTTGGAGCGTTCGTTTTCATG 5107  
Qy 5665 cacttcaactgtgagccgagctgctcatttctcctaaaaatcagagaatccgagtggtcc 5724  
Db 5108 CGCTTCATTGTGGTGTGGAGTGTGCATTTCTTAAATCCGAGAAATCAGGGTGGTCC 5167  
Qy 5725 tgggtgaaggttaagcagagagctgtgctatccagctccttacttggtgatgaatggag 5784  
Db 5168 TGGTGAAGAAAGCCAGAGGCTGTGCTTACCAGCCCTTACTTGGATGAATATGGAG 5227  
Qy 5785 aaacagacctggcctgaagaggggcaaccccttcttattctctgagcggtatcga 5844  
Db 5228 AAACAGACCCAGGCTAAAGAGAGGAAACCCACTTCAATTATCTCGGAGCGGTATCGGA 5287  
Qy 5845 agctccatttggctggaacacacatgctcattatagaagagattgctaggagccaagaga 5904  
Db 5288 AGCTGCATTTGCTGCGCAACAGCACTGCATTATAGAAGAGATTGCTCGGAGCCAGAGA 5347  
Qy 5905 ctaatcagatgtatttgattcaactgcagttactgagctccaactcgtcccaag 5964  
Db 5348 CTAATCAGATGTATTGTGATTGAGTTAGCTGCGAGTTACTCTGAGCTTCAGTTCTGCCCTCAAG 5407  
Qy 5965 acaatcaaatgacgacagtagtaaaaggctgattcaaatattggaactttctgag 6024  
Db 5408 ACAATCATGAGTACATCAATAAAGACTGATCTAAATCTAGAGAACCTTCTCGAGG 5467  
Qy 6025 gctgggaagtagttggaggtcttttgcctcattgctccaggttccacttacatcaataaat 6084

Db 5468 ACGGGGAAAGTATTGGAGGCTCTTTTGATCCATGTCCAGATTCCACACATTAATAAAAT 5527  
Qy 6085 atttcttaatgaggtattgcttttcaattagcaaaaacatactctcaagggaaaaa-aggac 6143  
Db 5528 ATTCTCTAATGGAATATTGCTTTTCAATTATCAAAACATAAGCTTCAAGGGAACAAAGAC 5587  
Qy 6144 atagatcaaatctgt-----ttt 6160  
Db 5588 ATAGATTAATGTTTTATGTCTTAGAACACTAAAGAAATGCTTGTTCATCCAAGTGTCTAT 5647  
Qy 6161 atgtgtagtatttccaggaaatttattccctctcaataattgtctcatttcttatttt 6220  
Db 5648 TTCTGCTAATATTTCAGAAACTCTCTTCCCTTCATAACTGCTCTAGTTTCATTTCATAT 5707  
Qy 6221 catccacttggtagtagaagtcacgtcaaacagcttgtagacattttatgcttgggttaac 6280  
Db 5708 CACCCACCTGGTTAATGAGGTACATTAAGCATTTTGTGGACATTTCTCCATCTGGCTAAC 5767  
Qy 6281 tcttctgcaatttattgttgggttt 6307  
Db 5768 ATCTCTGCACCTTTGTATTGTGGTGT 5794

RESULT 2  
US-09-228-317-1  
; Sequence 1: Application US/09228317  
; Patent No. 6159732  
; GENERAL INFORMATION:  
; APPLICANT: Varshavsky, Alexander  
; APPLICANT: Kwon, Yong Tae  
; TITLE OF INVENTION: NUCLEIC ACID ENCODING MAMMALIAN UBRL  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Kevin M. Farrell, P.C.  
; STREET: P.O. Box 999  
; CITY: York Harbor  
; STATE: ME  
; COUNTRY: US  
; ZIP: 03911  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/228,317  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Farrell, Kevin M.  
; REGISTRATION NUMBER: 35,505  
; REFERENCE/DOCKET NUMBER: CIT-2001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (207) 363-0558  
; TELEFAX: (207) 363-0528  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 6395 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 115..5385  
US-09-228-317-1

Query Match 70.3%; Score 4435,4; DB 3; Length 6395;  
Best Local Similarity 86.3%; Pred. No. 0;  
Matches 4997; Conservative 0; Mismatches 721; Indels 69; Gaps 6;  
Qy 590 ttcaggggcccgtcgttaaaagtgtgcctctgtctc-tccgaccggccacaggtttccgct 648



Qy 2809 ataatctctgttactgtacttcagaggtatgaacttgcagagcttttaacaagacca 2868  
Db 2228 ACAAGTCTTGTACTGTGACTTCAGAGATATGAATCTTACTGATGCTTTTAACAAGACCA 2287  
Qy 2869 tatctacaaaagaccagatttgattaaacaataaatacacactaataagaagaatgcttc 2928  
Db 2288 TATCCACAAAAGACCAGATTGTATTAACACAGTATAATACATTAATAGAGAATAATGCTTC 2347  
Qy 2929 aggtctcatctatatgttgggtgagcgttatgtacctggagtggaagtgtgacaaag 2988  
Db 2348 AGGTCTCTCATATATGTGGGAGAACGTTATGTACCTGGAGTGGGAAATGTTACACAG 2407  
Qy 2989 aagagtcacaatgagagaaatcattcacttgcgttttgatgaaccccatgcaacacagt 3048  
Db 2408 AGGAGGTTTATATGAGAGAGATTACTCTACTTCTTGTGATTGAGCCCATGCCACACAGTG 2467  
Qy 3049 ccattgccaataatttactgagaatgaaataatgaactggcttagagaatgtcataa 3108  
Db 2468 CCATGCCCAAAACCTACTTGAGACGAAATATGAATGAACTGGCTTAGAATAATGTCATAA 2527  
Qy 3109 acaaagtgccacatttaagaacaccagtggtatcaggccatggagtttatgaactaaaag 3168  
Db 2528 ACAAAGTGGCCACATTTAAGAAACACAGGTGTGTGGGCCATGGAGCTTATGAATTGAAG 2587  
Qy 3169 atgaatcactgaagacttcaatatgtacttttatacttccataaccccgcatagca 3228  
Db 2588 ATGAATCTACTGAAGACATTCATATGTACTTTTACCATTTATCTAAAAACACACGATAGCA 2647  
Qy 3229 aggtgaaacatatcagaagaaaaggagaaaacagaagaaacaaagatgaagcattgcgc 3288  
Db 2648 AGGCTGAACATATGCAAGAAGAAAGAGAAACAAAGAAATAAAGATGAAGCTTGCCGC 2707  
Qy 3289 caccaccctcctgaaattctgcctgttttcagcaaatgatttaaccttccaactgtg 3348  
Db 2708 GCCCACTCTCCAGAGTCTGCCCTGCTTTCAGCAAGATGCTCAACCTGCTCAGCTGTG 2767  
Qy 3349 atatcatgtatgtacattctcaggaccgttatgtgagcgggaatagacacagatttcaact 3408  
Db 2768 ATGTATGATATATACATCTCAGGACCATCTTTGAGCGGCAGTGGACACGGAGTCTAATC 2827  
Qy 3409 tgtgacgaagggatgtccaaatggcttttcataattctcgttcattggcattgggttactagaag 3468  
Db 2828 TGTGACAGAGAGGATGTGCAGATGGCGTCCATATATTGGCAGTGGCTTGTGGGAAG 2887  
Qy 3469 agaagcaacagcttcaaaaagctcctgaagaagaagtaacatttgacttttatcataag 3528  
Db 2888 AGAAGCAGCAGCTTCAGAAAGCTCCTGAAGAGAGAGTGGCTTTGACNTTTTACCATAAG 2947  
Qy 3529 cttaagattgggaagttcagccat-----gaatatacaaatgcttttggaaaaac 3579  
Db 2948 CTTCAGATTGGGAAGTTCAGCCATGAATGCTCAGAAATATACAAATGCTTTGGAAAGAC 3007  
Qy 3580 tcaaaaggaattcccccagttagaagccagaagaacatgaataacgttggtacttcagatgt 3639  
Db 3008 TCAAAAGGATCCCCAAATTAGAAGGCCAGAGGACATGATAACATGGATFACCTCCAGATG 3067  
Qy 3640 ttgacacagtgaagcatttaagaaaaatctgttttaattgtagcaaacacacatcaggt 3699  
Db 3068 TTGACACAGTGAAGCGATTAAAGAAAAATCTGTGTAGTTGGCAACCACTTCAGGAC 3127  
Qy 3700 cggaatctattagaatgatgatgattactcatgataaagaaaaagcagaaacaaaaagaa 3759  
Db 3128 TGGAGTGCAATTAAGAGTGAAGAGATTACTCATGATAAAGAAAGGCAGAGAACGAGAA 3187  
Qy 3760 aagctgaagctgtaggctacatgccagaagatcatggctcagatgtgctgcttacaaga 3819  
Db 3188 AAGCTTGAGGCCCTAGGCTTTCGCCAGAAGATCATGCCCCAGATGTCTGCTTACAGA 3247  
Qy 3820 aaaaattcaatgaaactcaaaactcatgtatgacaatacatcagaataatgcctgggaag 3879  
Db 3248 AAAACTTCATTTGAAACCCCAAAACTCATGTATGATAATACGTACAGAAGTAACAGGGAAG 3307

Qy 3880 aagattccattatggaggaagagcaccacagcagtcagtcagtcactactctagattgctt 3939  
Db 3308 AAGACTCCATTATGGAGGAAGAGAGACACCTFCAGCAGTCAGTGAGCCCTCTAGAAATGCTC 3367  
Qy 3940 tgggtctctaaacgggtcccatctgttactgaagaagaggtgctgacgtgcatcctctggcc 3999  
Db 3368 TGGGCCCTTAAACGGGGCCGGCTGTTACCGAAAAGAGAGGTGCTCAGCTGCATCCTCTGCC 3427  
Qy 4000 aagaagaacagaggtgaaaaataagaaaaataatgccattggtattatcgctgctgtgtccaga 4059  
Db 3428 AAGAAGAACAAGAGGTAAACCTAGAAAAATATGCCATGTTGTTCAAGCATGTGTGCAGA 3487  
Qy 4060 aatctactactcttaacccacgacacaggggaaacccatagaaactcctcaggagaagccctag 4119  
Db 3488 AATCCACCCGCCCTAACCCACGACACAGAGGGAAGCCTGTGGACCCTTATAGGGAAACACTGG 3547  
Qy 4120 accacttttcatgtatccagacttggcatatggaactttacagaggaagctgtggtcatg 4179  
Db 3548 ACCCTCTTTTCATGTGATCCAGACTTGGCATGGAACCTTATACAGGAAGCTGTGGTCTATG 3607  
Qy 4180 taatgcacgagtgctgtgcgagaagtattttgaagctgtacagctgagctctcagcagc 4239  
Db 3608 TAATGCATGCAGTGTGCTGCAGAAAGTATTTTGAAGCTGTGCAGCTGAGCTCGCAGCAGC 3667  
Qy 4240 gaattcatgttgaccttttgaacttggaaaagtggagaataatcttggccctcttggaaat 4299  
Db 3668 GCATTCACCTTAGACCTGTTTGACCTGGAGAGCGCGAGTAGCTACCTATGCCCTCTGCAAGT 3727  
Qy 4300 ctctgtgcaactctgtgaccccatattctcttggaaacctcaaaagataaaacagtga 4359  
Db 3728 CTCTCTGCAACACTGTATCCCCCATCATCCCTTTCGACGCCGACAGAAATCAACAGTGA 3787  
Qy 4360 atgcagatgctcttgcctcaacttttgacctggcagcgttgatcacagactgttctggtcca 4419  
Db 3788 ATCGGAGGCTCTTGCTCAACTTTTGACCTTGCCCGGTGGATACACACTGTCTCTGCCA 3847  
Qy 4420 gaatacaggttataataataagaacatgctaaagagaaaaacc---caatctctatttct 4476  
Db 3848 GAATATCGGTTTATAATATATAAGCATGTCTAAAGGAGAAAGCCCAAGCAGTAGTTCTCTGT 3907  
Qy 4477 ttaatacaagaaatggagattctacttggagttccattccactcctcagtttggcgtt 4536  
Db 3908 TTAATCAAGAAATGGGGGATTCAACTTTTGAGTTTCAATTCATCTCAGTTTTGGAGTTC 3967  
Qy 4537 agcttcctgattaaatattcaaatagcatcaagaaaaatggttattctcttcttggccaacaa 4596  
Db 3968 AGCTTTCGGTGAATAATTCAAAATAGTATCAAGGAAATGGTCAATTCCTCTGSCCAACAA 4027  
Qy 4597 ttatgaattgattgaaaagtgccacctgatgaaaagggatcctcaggtccccaatgtga 4656  
Db 4028 TTTTACAGAAATTTGGCTGAAAGTGCTCTCTGATGAACCTAGACCCACGAGTGCCCATGTGA 4087  
Qy 4657 cctggagacctgcgtcttccactatccaggaattgaaaatctattgggagatgaagaa 4716  
Db 4088 CTTGGAGACAGTGTGGCTTACCATCCAGGCAATTCGAAACCTGTTGGGAGATGAAGGAA 4147  
Qy 4717 aacctgttttgagcacttcaaaatgagcagcataatggtctgaagaagcataatgagct 4776  
Db 4148 AACCTCTATTGGAGCACTTCAAAATAGACAGCATACGCGTCTGAAGCGCTATATGCACT 4207  
Qy 4777 ttgcagttgcacagagattacctgtcctcaggtcctgtatagacagaaacatctgttctcgtc 4836  
Db 4208 TTSCAGTTGCACAGAGGGCTACCTGCCCTCAGGTCCTGTATACATACAAACATCTGGCTGGC 4267  
Qy 4837 ttctaactgtgttcttctaatacaataaaatcagaagatacaccatgctctctgtctatag 4896  
Db 4268 TCCTGTCAAGTTATCTTCTTAACCTGCAATCAGAAAAATACACCAGGCTTCTGTCTGTGG 4327  
Qy 4897 atctgtttcaatgtttgggtgggtgctgttagcattcccatcctcttatgtggatgacc 4956  
Db 4328 ATCTCTTCCATGTTCTGGTGGCGCAGCTTAGGCTTCCCATCTCTGTATTGTTGGGATGACA 4387  
Qy 4957 ctgtgatctgcagcctcttctcagtttagttcttcttaataaccaccttattctcttccatt 5016



Db	4388	CCGTGGATCTGCAGCGCTCGCCACTTAGTCTTCATATAACCAACCTCTACTCTCTCCATC	4447
Qy	5017	tgatcacattggcacacatgcttcagatactactacttcaagtagaca-----cag	5064
Db	4448	TGATCACCATGCGCACATGCTTCAGATACTCTTTACACAGATACAGATCTGTCTCCAG	4507
Qy	5065	gcttacccttctcaggttcagaagaacagtgaaagagctcatctccgacatcttcttct	5124
Db	4508	GGCCGCCGCTTCTGAGGGTGAAGAGGATAGTGAAGAGGCTCGCTGTGATCTGCTTTCT	4567
Qy	5125	ttgcagaaatttctcaatatacaagtggctccattgggtgtgatatctctggctggtatt	5184
Db	4568	TTGTGGAAGTGTGCGAGCACACAGACGGGCTCACTGGGTGCGGTGTCTCCGGCTGTACC	4627
Qy	5185	tgtgggttctactgaagaataggcctaccccccttatcttcgtgtgtcgtcattgttttcc	5244
Db	4628	TGTGCTCTCTCCCTGAGAACGCGCATCACCCCTTACCTCCGCTGTGTGCACTGCTTTTC	4687
Qy	5245	actatttacttgggttaactccgcctgaggaactgcataccaaattctgcagaaggaagt	5304
Db	4698	ACTATTACTTTGAGTAGTCCCGCTGAAGAACTGTGTGCCAAATCTGCTGAAGGAGAAT	4747
Qy	5305	acagtgcactctgtagctatctatttaccacaaaattgttcctgctctctccaggaat	5364
Db	4748	TCAGTGCACCTGTAGCTATCTATTACCCACAAATTTGTTCTGCTCTTCCAGGAAT	4807
Qy	5365	attgggatactgtgaaggcccttgcctccagaggtgggtgtgcagatcctgccttactaaact	5424
Db	4808	ATTGGGATACCAATAAGGCCCTTACTACAGAGGTGGTGTGGAGATCCTGCTTACTCAAGT	4867
Qy	5425	gtttgaagcaaaaaaacaccgtggtcaggtagccctagaaaaaagaatagtttgatagagc	5484
Db	4868	CTTTGAAGCAGAAAAGTGTGTGGTAGGTACCTAGAAAAAGAAATAGTTTGTATAGAGC	4927
Qy	5485	ttcctgatgactatagctgcctccctgaaatcaagctctctcatctcaggtgccacagctcg	5544
Db	4928	TTCTTGAGGACTACAGCTGTCTTCTAATCAGGCTTCTCACATTTAGTGTCCACGGTCTG	4987
Qy	5545	cagatgatgaggaagcatcctgctcctctgcctttctgtggggctatactatgtcttc	5604
Db	4988	CAGATGATGAGCGAAAGCATCTGTCTCTGTCTTTCTGTGGGGCCATCCTGTGTTCTC	5047
Qy	5605	agaacatttgcgcaggaaaattgtgaacgggggaagaggttgagctttgcattttccag	5664
Db	5048	AGAACATCTGTTGCCAAGAAATAGTGAATGGGGAAGAGGTTGGAGCGTGCCTTTTTCATG	5107
Qy	5665	cacttcactgtgagcgagctcgcatttctcaaaatcagaaatgccagtggtctcc	5724
Db	5108	CGCTTCATGTGGTCTGGAGTCTGCATTTTCTTAAATCCGAAATCCAGGGTGGTCC	5167
Qy	5725	tggttgaaggtaaagccagaggtgctcctatccagctccttacttggatgaatactggag	5784
Db	5168	TGGTGGAAAGGAAAAGCCAGAGCTGTGCTACCCAGCCCTTACTTTGGATGAATATGGAG	5227
Qy	5785	aaacagaaccttggcctgaagaggggcaaccccccttoatttatctcgtgagcgggtatcga	5844
Db	5228	AAACAGACCCAGGGCTAAAGAGAGAAACCCACTTCATTATCTCGGGAGCGGTATCGGA	5287
Qy	5845	agctccatttgctcggcaacaacactccattatagaagagattgctaggagccaagaga	5904
Db	5288	AGCTGCATTTGTCTGGCAACAGCACTGCATTTATAGAAGAGATTGCTCGGAGCCAGGAGA	5347
Qy	5905	ctaatcagatgttatttggattccaactggcagttactgtgagctccaactctgcctccaag	5964
Db	5348	CTAATCAGATGCTATTTGGATTTAACTGGCAGTTACTCTGACTTCAGTTCTTGCCCTCAAG	5407
Qy	5965	acaatcaaaatgacgacagtagtaaaaggtgattcaaaattatggaaaactttctgag	6024
Db	5408	ACAATCATGAGTGACATCAATAATAAAGACTTGATCTAAATTTCTAGAGAACTTTCTGAGG	5467
Qy	6025	gctgggaagatttggagggtcttttggctccatgtccaggttcaacttcatcataataaat	6084

Db	5468	ACGGGGGAAGTATTGGAGGGTCTTTTGGATCCATGTCAGATTCCACACATTATATAAAAT	5527
Qy	6085	attctctaagagagattgctttccaattagcaaacataatgcttcacaggaaaaa-aggac	6143
Db	5528	ATTCTCTTAATGGAATATTGCTTTCAATTTTCAAAACATAAGCTTCAAGGGGAAAAACAAGAC	5587
Qy	6144	atagatcaatctgt-----ttt	6160
Db	5588	ATAGATTAAAGTTTATGTTCTAGAACACTAAAGAAATGCTTTGTTCAACAAGTGCTCAT	5647
Qy	6161	atgtgctagatttccaggaaatttattcccttccataaattgtctcattcttatttt	6220
Db	5648	TTCTGCTAATATTTCAGAAAACTCCTTTCCCTTCATAACTGTCCTAGTTCAATTTTCATAT	5707
Qy	6221	catccacttggtagatgaagtcacgtcaaacagttgtagacattttatggtgggttaac	6280
Db	5708	CACCCACCTGGTAAATGAGGTGCACATTAAGCATTTGTGGACATTTCTCCATCTGGGCTAAC	5767
Qy	6281	tcttctgcattttgtatttgggtgttt	6307
Db	5768	ATCTCTGCACCTTTGTATTGGTGTTT	5794
RESULT 3			
US-08-982-956-2			
; Sequence 2, Application US/08982956			
; Patent No. 5861312			
; GENERAL INFORMATION:			
; APPLICANT: Varshavsky, Alexander			
; APPLICANT: Kwon, Yong Tae			
; TITLE OF INVENTION: NUCLEIC ACID ENCODING MAMMALIAN UBR1			
; NUMBER OF SEQUENCES: 2			
; CORRESPONDENCE ADDRESS:			
; ADDRESSEE: Kevin M. Farrell, P.C.			
; STREET: P.O. Box 999			
; CITY: York Harbor			
; STATE: ME			
; COUNTRY: US			
; ZIP: 03911			
; COMPUTER READABLE FORM:			
; MEDIUM TYPE: Floppy disk			
; COMPUTER: IBM PC compatible			
; OPERATING SYSTEM: PC-DOS/MS-DOS			
; SOFTWARE: PatentIn Release #1.0, Version #1.25			
; CURRENT APPLICATION DATA:			
; APPLICATION NUMBER: US/08/982,956			
; FILING DATE:			
; CLASSIFICATION:			
; ATTORNEY/AGENT INFORMATION:			
; NAME: Farrell, Kevin M.			
; REGISTRATION NUMBER: 35,505			
; REFERENCE/DOCKET NUMBER: CIT-2001			
; TELECOMMUNICATION INFORMATION:			
; TELEPHONE: (207) 363-0558			
; TELEFAX: (207) 363-0528			
; INFORMATION FOR SEQ ID NO: 2:			
; SEQUENCE CHARACTERISTICS:			
; LENGTH: 1001 base pairs			
; TYPE: nucleic acid			
; STRANDEDNESS: single			
; TOPOLOGY: linear			
; MOLECULE TYPE: cDNA			
; FEATURE:			
; NAME/KEY: CDS			
; LOCATION: 1..999			
US-08-982-956-2			

Query Match	15.8%	Score 996.2	DB 2	Length 1001
Best Local Similarity	99.7%	Pred. No. 7.6e-279		
Matches 998	Conservative 0	Mismatches 3	Indels 0	Gaps 0
Qv	2799	atgatcccaataaagttcttcttacttgatcttcagaagatgaacttccgaagcctttt	2858	

```
Db 1 ATGGATCCCAACAGTCTTGTTACTGTTACTGATGAGATGAACTGCCAGGCTTTT 60
QY 2859 acaagaccatctacaaaacacagagattgattataacaataataacactaatagaa 2918
Db 61 AACAGACCATATCTACAAAACACAGGATTTGATTAAACAATAATAACACTAATAGAA 120
QY 2919 gaaatgcttcaggtcctcatctatatttggtgagcgcttattgacccgagtggaat 2978
Db 121 GAAATGCTTCAGTCTCATCTATATTGTGGTGAGCGTTATGATACCTGGAGTGGGAAT 180
QY 2979 gtgaccaagaaggtcacatgagagaatacctcactcttcttgattgaacccatg 3038
Db 181 GTGACCAAGAAGAGGTCAATGAGAGAAATCATTCACCTTGCTTGTGCAATGAACCCATG 240
QY 3039 ccacacagtgccatgcccacaaattaccctgagatgaaataaataaataaactgcttagag 3098
Db 241 CCACACAGTGCCATGTCACAAATTTACCTGAGATGAAATTAATGAACACTGGCTTAGAG 300
QY 3099 aatgtcataaacaagtgccacatttaccctgagatgaaataaataaactgcttagag 3158
Db 301 AATGTCTATAAACAAGTGCCACATTTAAGAAACACAGGTGTATCAGGCCATGGAGTTTAT 360
QY 3159 gaactaaagatgaatcactgaaagacttcaatattgacttttatactactccaaaacc 3218
Db 361 GAACTAAAGATGAATCACTGAAGACCTCAATATGTACTTTTATCATTTACTTCCAAACC 420
QY 3219 cagcatagcaggtgaacatgacagaagaagaagaagaagaagaagaagaagaatgaa 3278
Db 421 CAGCATAGCAGGCTGAACATATGACAGAAAGAGAGAAACAAAGAAACAAAGATGAA 480
QY 3279 gattgcccgcaccacccctcctgaattctgcctgctttcagcaaatgataacctt 3338
Db 481 GCATTTGCCGCCACACACCTCTCTGAATTCGCTGCTTTTCAGCAAGTATTAACCTT 540
QY 3339 ctcaactgtgatatacatgatgtacattctcaggaacgcttatttgagcgggcaatagacaa 3398
Db 541 CTCAACTGTGATATCATGATGATGATCTCTCAGGACCGTATTTGAGCGGCAATAAACACA 600
QY 3399 gatttcaactgtgacgaagaggtgctccaaatggcttttcatattctgcttgcttgggt 3458
Db 601 GATTTCTAACTTTGTGACCGAAGGAGTCTCCAAATGGCTTTTTCATATTCTGCGATTGGGT 660
QY 3459 ttactagaagaagaacacagcttcaaaagctccctgaagaagaagtaacatttgacttt 3518
Db 661 TTACTAGAGAGACACACAGCTTCAAAAGCTCTCTGAGAGAGAGTAAACATTTGACTTT 720
QY 3519 tatcataaggttcaagattgggaagttcagccatgaaataatacaaatgcttttggaaaaa 3578
Db 721 TATCATAGGCTTCAAGATTGGGAAGTTCAGCCATGAATATACAAATGCTTTTGGAAAAA 780
QY 3579 ctcaagaagaattccccagttagaagccagaagggacatgataacgtggatactcagatg 3638
Db 781 CTCAAAGGAATTTCCCAAGTTAGAGGCCAGAGGACATGATAACGTGGATCTTCAGATG 840
QY 3639 ttgacacagtgaaagcgttaagagaataatctgtttaattgtagcaaccacatcaga 3698
Db 841 TTTGACACAGTAGAGCGATTAGAGAAAAATCTTGTTTAATTGTAGCAACCCATCAGGA 900
QY 3699 tcggaatcttataagaatgatgattactcatgataaagaagaagaagaagaagaagaaga 3758
Db 901 TCGGAATCTATTAAAGATGATGAGATTACTCATGATAAAGAAAAAGACAGACGAAAAAGA 960
QY 3759 aaagctgaagctgctaggtacatgcagcagaagatcggc 3799
Db 961 AAAGCTGAAGCTGCTAGGCTTTCGCCCAAGAGATCATGGC 1001
```

## RESULT 4

US-09-228-317-2

; Sequence 2, Application US/09228317

; Patent No. 6159732

; GENERAL INFORMATION:

```
APPLICANT: Varshavsky, Alexander
TITLE OF INVENTION: NUCLEIC ACID ENCODING MAMMALIAN UBRL
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Kevin M. Farrell, P.C.
STREET: P.O. Box 999
CITY: York Harbor
STATE: ME
COUNTRY: US
ZIP: 03911
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/228,317
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Farrell, Kevin M.
REGISTRATION NUMBER: 35,505
REFERENCE/DOCKET NUMBER: CIT-2001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (207) 363-0558
TELEFAX: (207) 363-0528
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1001 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..999
US-09-228-317-2
```

```
Query Match 15.8%; Score 996.2; DB 3; Length 1001;
Best Local Similarity 99.7%; Pred. No. 7.6e-279;
Matches 998; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2799 atgattcccaataagttctgttactgttactcagaggtatgaacttgcgaggtttt 2858
Db 1 ATGGATCCCAACAGTCTTGTTACTGTTACTGATGAGATGAACTGCCAGGCTTTT 60
QY 2859 acaagaccatctacaaaacacagagatttgaataacaataataacactaatagaa 2918
Db 61 AACAGACCATATCTACAAAACACAGGATTTGATTAAACAATAATAACACTAATAGAA 120
QY 2919 gaaatgcttcaggtcctcatctatatttggtgagcgcttattgacccgagtggaat 2978
Db 121 GAAATGCTTCAGTCTCATCTATATTGTGGTGAGCGTTATGATACCTGGAGTGGGAAT 180
QY 2979 gtgaccaagaaggtcacatgagagaatacctcactcttcttgattgaacccatg 3038
Db 181 GTGACCAAGAAGAGGTCAATGAGAGAAATCATTCACCTTGCTTGTGCAATGAACCCATG 240
QY 3039 ccacacagtgccatgcccacaaattaccctgagatgaaataaataaactgcttagag 3098
Db 241 CCACACAGTGCCATGTCACAAATTTACCTGAGATGAAATTAATGAACACTGGCTTAGAG 300
QY 3099 aatgtcataaacaagtgccacatttaccctgagatgaaataaactgcttagag 3158
Db 301 AATGTCTATAAACAAGTGCCACATTTAAGAAACACAGGTGTATCAGGCGCATGGAGTTTAT 360
QY 3159 gaactaaagatgaatcactgaaagacttcaatattgacttttatactactccaaaacc 3218
Db 361 GAACTAAAGATGAATCACTGAAGACCTTCAATATGTACTTTTATCATTTACTTCCAAACC 420
QY 3219 cagcatagcaggtgaacatgacagaagaagaagaagaagaagaagaagaagaatgaa 3278
```

```
|||||
Db 421 CAGCATAGCAGGCTGACATATCCAGAGAAAGGAGAAACAAAGAAACAAAGATGAA 480
QY 3279 gcattgcgcacacacacccctcctgaattctgcctgctttcagcaaaagtattaaacctt 3338
Db 481 GCATTGCGGCACACACACCTCTCTGAATCTGCGCTGCTTTTCAGCAAAAGTATTAACTT 540
QY 3339 ctcaactgtatcatgatgatcatctcagagccgtatttgagcgggcaatagacaca 3398
Db 541 CTCNACTGTGATATCATGATGATACATCTCTCAGAGCCGCTATTGAGCGGGCAATAACACA 600
QY 3399 gattctaactgtgaccggaaggatgctccaaatggctttcatattctgcattgggt 3458
Db 601 GATTCTAATCTGTGACCGAAGGATGCTCCAATGCGCTTTTCATATCTCGCATTTGGGT 660
QY 3459 ttactagaagaagcagcacagcttcaaaaagctcctgaagaagtaacatttgacttt 3518
Db 661 TTACTAGAAGAAGAACACACAGCTTCAAAAAGCTCCTGAAGAAGTAACATTTGACTTT 720
QY 3519 tatcataaggctccaagattgggaagttcagccatgaatatatacaaaatgcttttggaaaaa 3578
Db 721 TATCATGAAGCTTCAAGATTGGGAAGTTCAGCCATGAATATACAAATGCTTTTGGAAAAA 780
QY 3579 ctcaaaaggaattcccccagttagagggccagaagacatgataacgtgatacttcagatg 3638
Db 781 CTAAAGCAATTCCTCAGTTAGAGGCCAGGAAGACATGATTAACGTGGATCTTCAGATG 840
QY 3639 ttgacacagtgaaagcagttaaagagaaaaatctgtttaattgtagcaaacacatcagga 3698
Db 841 TTTGACACAGTGAAGCGATTAAAGAGAAAAATCTGTTTAATGTTAGCAACCATCAGGA 900
QY 3699 tcggaatcttataagaatgatgagattactactatgaataaagaaagcagaaacgaaaaa 3758
Db 901 TCGGAATCTATTAGAATGATGAGATTACTATCATGATAAGAAAAAGCAGAACGAAAAAGA 960
QY 3759 aaagctgaagctagctacatcgccagaaagatcgtgc 3799
Db 961 AAAGCTGAAGCTGTAGGCTTCATCGCCAGGAAGATCATGGC 1001

RESULT 5
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
```

```
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTgpt-F1s
US-08-232-463-14

Query Match 0.9%; Score 57.4; DB 1; Length 7218;
Best Local Similarity 9.4%; Pred. NO. 6e-06;
Matches 43; Conservative 219; Mismatches 195; Indels 0; Gaps 0;

QY 2862 aagaccatatctacaaaagaccaggatttgattaaacaataataacactaataagaa 2921
Db 1481 AATTACCTATCTATCGAAGTAGTTAAAGAGATAGAGAATTTGCTACRRRRRRRRRR 1422
QY 2922 atgcttcaggtcctctatctatgttggtgagcgttatgtacctggagtggaatgtg 2981
Db 1421 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1362
QY 2982 accaaaagagaggtcacaaatgagagaaaaatcattcactgcttcttgattgaacccatgcca 3041
Db 1361 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1302
QY 3042 cacagtgcctatgcacaaaatttacctgagaaatgaaataaataagaaactgcttagagaaat 3101
Db 1301 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1242
QY 3102 gtcatataaacaagtggccacatttaagaacccagggtgtatcgatggccatgagtttatgaa 3161
Db 1241 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1182
QY 3162 ctaaaagatgaatcactgaaagacttcaatatgtacttttatctacttcccaaaacccag 3221
Db 1181 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1122
QY 3222 catagcaaggctgaacatcatgagaaagaaagaaagaaagaaagaaagaaagaaagatgaagca 3281
Db 1121 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1062
QY 3282 ttgcccgcaccaccctcctcctgaatttctgcctgctt 3318
Db 1061 AAGCTCCCTCGACCTCGACCGACCAAGTCGGAATTAATT 1025

RESULT 6
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```



```
;
; TELEFAX: 650-327-3231
;
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 72928 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
;
US-09-009-913-1

Query Match          0.7%; Score 44.2; DB 3; Length 72928;
Best Local Similarity 86.0%; Pred. No. 0.18;
Matches 49; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 303 cctgtgaaacacacattctctgtctgtatgaattgactactctagctgag 359
|||||
Db 19091 CCCGGCAACACACATCTACTTCTCTCTATGATTTGACCACCTCTAGTAGCT 19035

RESULT 9
US-08-991-789A-236/c
; Sequence 236, Application US/08991789A
; Patent No. 6225054
;
GENERAL INFORMATION:
; APPLICANT: Frudakis, Tony N.
; Smith, John M.
; Reed, Steven G.
;
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
TREATMENT AND DIAGNOSIS OF BREAST CANCER

NUMBER OF SEQUENCES: 292
CORRESPONDENCE ADDRESS:
ADDRESSEE: Seed IP Law Group
STREET: 701 Fifth Avenue, Suite 6300
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
;
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/991.789A
FILING DATE: 11-Dec-1997
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Potter, Jane E. R.
REGISTRATION NUMBER: 33,332
REFERENCE/DOCKET NUMBER: 210121.419C3
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 236:
SEQUENCE CHARACTERISTICS:
LENGTH: 325 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
;
SEQUENCE DESCRIPTION: SEQ ID NO: 236:
US-08-991-789A-236

Query Match          0.7%; Score 42.2; DB 4; Length 325;
Best Local Similarity 79.4%; Pred. No. 0.021;
Matches 50; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 298 acacacccctggaacacacattctctgtctgtatgaattgactactctagctgg 357
```

```
Db 283 AAAAGCCCTGGCAACACACATCTATTGTGTTCTATGAATTTGACTACTTAGGTAC 224
QY 358 atc 360
||
Db 223 CTC 221

RESULT 10
US-08-963-602-6/c
; Sequence 6, Application US/08963602
; Patent No. 6090554
;
GENERAL INFORMATION:
; APPLICANT: Woychik, Richard
; Garfinkel, David
;
TITLE OF INVENTION: EFFICIENT CONSTRUCTION OF GENE
TARGETING VECTORS
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
STREET: 233 South Wacker Drive/6300 Sears Tower
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
;
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/963,602
FILING DATE:
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Clough, David W.
REGISTRATION NUMBER: 36,107
REFERENCE/DOCKET NUMBER: 01017/33985
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 474-6300
TELEFAX: (312) 474-0448
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1517 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "primer"
US-08-963-602-6

Query Match          0.6%; Score 41; DB 3; Length 1517;
Best Local Similarity 56.2%; Pred. No. 0.13;
Matches 77; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 109 gagaacttaacctcaaggcttatctgtcttccacttatccaaagaagctgtatgc 168
|||||
Db 1036 GTGTCCCTCGCTCAGAAAGCCGTCTCTTTGGGGCGCTGGTAGCCGGCTGCTGCTGC 977
|||||
QY 169 cactgctgtgcccgcacctggaaccacccagccccactactgctccactaccactgg 228
|||||
Db 976 AGTCCCGCTGCCGCGCTGCTGCCACCACCAACATTGCTACTGCGCGCTCCACCACTGC 917
|||||
QY 229 ttctcccaccctgatca 245
|||
Db 916 TGCCTCCTCCTCCACCA 900

RESULT 11
US-09-128-155-16
; Sequence 16, Application US/09128155
; Patent No. 6117654
```

```

US-08-299-953-1
Query Match 0.6%; Score 39.8; DB 1; Length 2861;
Best Local Similarity 46.5%; Pred. No. 0.42;
Matches 128; Conservative 0; Mismatches 147; Indels 0; Gaps 0;

QY 3607 agaagagcatgataacgtgatacttcagatctgtttgacacagtcgaagcgattgaagaaaa 3666
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1290 AAAAAGAACGAGGAGCTTAATATATTTTGTAGATTTTACACGCTATTAAAAAAATATATCAA 1231
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3667 aatttggttaattgttagcacacatcagatcggaatctattataagaatgatgagatta 3726
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1230 AAAAATATCTTTATATATTAATAAATGGAGAGATAAATTTATATAATTTAAAAAAAAG 1171
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3727 ctcatgataaagaaaaagcagaaacgaaaaagaaagctgaagctgctaggctacatcgcc 3786
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1170 ATATAATAAATCTAGAGTTATATAAATAAACTAATATTAATTCUTTTTACATTTGTA 1111
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3787 agaagatcatggctcagatctgccttacagaaaaaacttcattgaaactcataaactca 3846
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1110 AAATGATTTATATATGATATAATTTTTTTTCAAAACAACCAATAATAAAAAATGATAGGG 1051
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3847 tgatgacatacatcagaaatgccctgggaaagaa 3881
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1050 AGTATTATCATATGTCAGAAATATTATTAAGAA 1016
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 13
US-08-459-415-1/c
; Sequence 1, Application US/08459415
; Patent No. 5744334
; GENERAL INFORMATION:
; APPLICANT: Dobres, Michael S. and Mandaci, Sevnur
; TITLE OF INVENTION: A Plant Promoter Useful for Directing the
; TITLE OF INVENTION: Expression of Foreign Proteins to the Plant Epidermis
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5744334ris
; STREET: One Liberty Place 46th. Floor
; CITY: Philadelphia
; STATE: PA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/459,415
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/299,953
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Beardell, Lori Y.
; REGISTRATION NUMBER: 34,293
; REFERENCE/DOCKET NUMBER: NOVA-0003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-564-8960
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2861 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-459-415-1

```

	Query Match	0.69;	Score 39.8;	DB 1;	Length 2861;
	Best Local Similarity	46.5%;	Pred. No. 0.42;		
	Matches 128;	Conservative 0;	Mismatches 147;	Indels 0;	Gaps 0;
QY	3607	agaaggacatgataacgtggatacttcagatcgtttgacacacagtgtaagcgattaaagagaaa	3666		
Db	1290	AAAAGAAGACGGAGGAGTAAATATTTTAGATTTTACACGATATTAATAAAATTTATATCAA	1231		
QY	3667	aatcttgtttaattgtagcaaccacatcagatcggaaatctcttaagaatatgagattta	3726		
Db	1230	AAAAATATCTTTATATATTAATAAGGAGAGATAAATTTATATAATTAATAAAAAAAG	1171		
QY	3727	ctcatgataaagaaaaagcagaaagcaaaagcgtgaagctgctaggtacatcgtcc	3786		
Db	1170	ATAATATAAAATCTAGAGTTATAATAAANAACATAATTAATCTTTTTTAACATTGTA	1111		
QY	3787	agaagatcatggtcgcagatgtctgccttacagaaaacttcattgaaactcataaactca	3846		
Db	1110	AAATGATTTATATATGATATAATTTTTTTTCAAAACACCACTAATAAAAAATGATAGG	1051		
QY	3847	tgtatgacaatacatcagaaatgcctggggaagaa	3881		
Db	1050	AGTATTATCATATGTCAGAAATATTATATAAGAA	1016		

```

RESULT 14
PCT-US95-11231-1/c
; Sequence 1, Application PC/TUS9511231
; GENERAL INFORMATION:
; APPLICANT: Dobres, Michael S. and Mandaci, Sevnuur
; TITLE OF INVENTION: A Plant Promoter Useful for Directing the Expression
; TITLE OF INVENTION: of Foreign Proteins to the Plant Epidermis
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & Norris
; STREET: One Liberty Place 46th. Floor
; CITY: Philadelphia
; STATE: PA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/11231
; FILING DATE: Herewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/299, 953
; FILING DATE: September 2, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Beardell, Lori Y.
; REGISTRATION NUMBER: 34,293
; REFERENCE/DOCKET NUMBER: NOVA-0016
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-564-8960
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2861 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
;
PCT-US95-11231-1

```

Query Match	0.6%	Score 39.8;	DB 5;	Length 2861;
Best Local Similarity	46.5%	Pred. No. 0.42;		

```

Matches 128; Conservative 0; Mismatches 147; Indels 0; Gaps 0;

Qy 3607 agaaggacatataacgtggacttcagatgtttgacacagtgtaagcgattaagagaaa 3666
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1290 AAAAAAGACGAGGGAGTAAATATTTAGATTTTACACGTAATAAAAAAATTTATACAA 1231
Qy 3667 aatctgttttaattgttagcaccacatcagatcggaaatctattaagaagtatgagatta 3726
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1230 AAAAATAATCTTTATATATATAAATAATGGAGAAGATAAATTTATATATAATTAATAA 1171
Qy 3727 ctcatgataaagaaaaagcagaaacgaaagaaagctgaagctgctaggctacatcgcc 3786
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1170 ATAATAATAAAATCTTAGAGTTAATAATAAAACCTAATATTAATTCITTTTAAACATTGTA 1111
Qy 3787 agaagatcgtgcagatgtctgctctacagaaaaactctcatggaactcataaactca 3846
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1110 AAATGATTTATATATGATATAAATTTTTTCAAAACACCAATATAAAAAATGATAGGG 1051
Qy 3847 tgtatgaacatcacatcagaaatgcctgggaaagaa 3881
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1050 AGTATTATCATATGTCAGAAATTTATATAAAGAA 1016

RESULT 15
US-08-299-953-2/c
; Sequence 2, Application US/08299953
; Patent No. 5646333
; GENERAL INFORMATION:
; APPLICANT: Dobres, Michael S. and Mandaci, Sevnur
; TITLE OF INVENTION: A Plant Promoter Useful for Directing the
; TITLE OF INVENTION: Expression of Foreign Proteins to the Plant Epidermis
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5646333ris
; STREET: One Liberty Place 46th. Floor
; CITY: Philadelphia
; STATE: PA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/299,953
; FILING DATE: Herewith
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Beardell, Lori Y.
; REGISTRATION NUMBER: 34,293
; REFERENCE/DOCKET NUMBER: NOVA-0003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-564-8960
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3881 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-299-953-2

```

Query Match	0.58;	Score 39.8;	DB 1;	Length 3881;
Best Local Similarity	46.5%;	Pred. No. 0.51;		
Matches 128;	Conservative	0;	Mismatches 147;	Indels 0;
Gaps	0;			
Qy	3607	agaaggacatgatcaactgatacttcagatgtttgacacagtcgaagcgtataagaaa	3666	
Db	1290	AAAAGAAGCAGGGAGGTATATTTTATGATTTTACACGTATTAAAAAATTTATATCAA	1231	





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 1, 2002, 12:32:35 ; Search time 2649.68 seconds  
(without alignments)  
25582.097 Million cell updates/sec

Title: US-09-724-126A-1

Perfect score: 6308

Sequence: 1 gccagaattgcgcagagg.....aatttgatttggtgtttt 6308

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```
EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_hic:*
10: gb_est1:*
11: gb_est2:*
12: gb_hic:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vrt:*
21: em_gss_other:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	674.6	10.7	682	10	AI929033
2	656.6	10.4	756	11	BG534574
3	637.4	10.1	641	10	AI361043
4	627.2	9.9	797	11	BG862813
5	573.2	9.1	782	11	BI086469
6	557	8.8	565	11	BF063405
7	512.2	8.1	606	10	BE589438
8	479.8	7.6	624	11	BG219270
9	477.8	7.6	481	10	AI192195
10	473.4	7.5	583	10	AW971391
11	471.6	7.5	478	10	AA401319
12	454	7.2	522	10	AW291190

13	449	7.1	522	11	BG382648
14	448	7.1	520	11	BG382624
c 15	410.8	6.5	486	10	BE650873
16	354.8	5.6	925	10	BE573578
c 17	352	5.6	374	10	AA400279
18	351.4	5.6	435	10	AW489271
19	330	5.2	394	11	BE774117
c 20	318	5.0	481	10	AI187306
21	307.2	4.9	397	10	AW311960
22	301.8	4.8	504	10	AI980640
23	266.2	4.2	327	10	BE654236
24	265.2	4.2	523	11	BG797647
c 25	263.4	4.2	300	11	Z17892
26	252.6	4.0	1079	11	BG292980
c 27	249	3.9	562	10	AI504731
28	243.6	3.9	372	11	BF542537
29	242.8	3.8	505	11	BG364916
30	241	3.8	325	10	AI615529
31	237	3.8	291	10	AV225341
c 32	235.8	3.7	396	11	BE930958
33	232.8	3.7	256	10	AI921294
c 34	231.4	3.7	278	10	BE077143
35	230.6	3.7	323	11	H33916
c 36	223.6	3.5	745	10	AW976158
37	222	3.5	555	11	BE873236
38	217	3.4	685	10	AI693180
39	209	3.3	313	10	AV168252
40	208.6	3.3	632	11	BG625558
41	206.8	3.3	221	10	AA507138
42	204	3.2	1148	11	BF164318
c 43	201	3.2	620	10	AW702134
44	200.4	3.2	435	11	BE930879
c 45	200	3.2	711	10	AI646734

#### ALIGNMENTS

RESULT 1

AI929033 682 bp mRNA EST 23-AUG-1999  
LOCUS au64c10.y1 Schneider fetal brain 00004 Homo sapiens cDNA clone  
DEFINITION IMAGE:2519538 5' similar to TR:070481 O70481 UBIQUITIN-PROTEIN  
LIGASE E3 COMPONENT N-RECOGNIN ;, mRNA sequence.  
ACCESSION AI929033  
VERSION AI929033.1 GI:5664997  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 682)  
AUTHORS Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,  
Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin  
J., Moore, B., Scheilberg, K., Steptoe, M., Tan, F., Theising, B.,  
White, Y., Wylie, T., Waterston, R. and Wilson, R.  
TITLE WashU-NCI human EST Project  
JOURNAL Unpublished (1997)  
COMMENT Other ESTs: au64c10.x1  
Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
This clone is available royalty-free through LLNL ; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Possible reversed clone: similarity on wrong strand  
Seq primer: -40RP from Gibco  
High quality sequence stop: 450.  
Location/Qualifiers  
1. 682  
/organism="Homo sapiens"

FEATURES  
source

```
/db_xref="taxon:9606"
/clone="IMAGE:2519538"
/clone_lib="Schneider fetal brain 00004"
/sex="male"
/tissue_type="frontal lobe"
/dev_stage="5 months post-conception"
/lab_host="DH10B"
/notes="Organ: brain; Vector: pBluescript SK (Stratagene);
Site_1: SstI; Site_2: XhoI; Double-stranded cDNA was
prepared from human fetal brain tissue. 5' and 3'
adaptors were used in cloning as follows: 5' adaptor
sequence:
5'-CAGAGAGAGAGAGCTCAAGGATCCTTAATTAATAATTCCTCCCTCCCTCC-3'
and 3' adaptor sequence:
5'-GAGAGAGAGAGCTGAGTCTTTTCTTTTCTTTT-3'. The library was
size-selected for >0.5 kb inserts and has an average
insert size estimated at 1.2 kb. This library was
constructed using the CAP-trapper method for full-length
enrichment and has not undergone amplification. Library
was constructed by Dr. Claudio Schneider (LNCIB-Area
Science Park, Trieste, Italy). "
1 others
BASE COUNT 178 a 155 c 166 g 182 t
ORIGIN

Query Match 10.7%; Score 674.6; DB 10; Length 682;
Best Local Similarity 99.3%; Pred. No. 3.9e-147;
Matches 677; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 5261 aactccgctcaggaactgcataccaattctgcagaagagagtagtcactctgtag 5320
Db 1 AACTCCGCTCAGGAACCTGATACCAATTCGCAAGAGAGAGTAGTCGACTCTGTAG 60

QY 5321 ctatctatttcacacataaattgtctctgtctctccagggaattgtggatactgtaag 5380
Db 61 CTATCTATCTTACCTACAAATTTGTTCTCTCTCCAGGAATATTTGGATACGTAG 120

QY 5381 gccctgtccagagtggtgtgcagatcctgcctactataaactgtttgaagcaaaaaa 5440
Db 121 GCCCTGTGTCAGAGGTGTGTGCAGATCCTGCCCTACTAAACTGTTTGAAGCAAAAAA 180

QY 5441 caccgtgtcaggtaccctagaaaaaataattgttagatagacttcctgactatag 5500
Db 181 CACCCTGTGTCAGGTACCTCTAGAAAAAGAAATAGTTGTATAGAGCTTCCGTGACTATAG 240

QY 5501 ctgctcctgaatcaagctctcattccagtgccacggtctgcagatgatgagcgaaa 5560
Db 241 CTGCTCTCTGAATCAAGCTTCTCATTTTCAAGTGCCACGCTGTGCAGATGATGAGCGAAA 300

QY 5561 gcatcctgtcctctgcctttctgtggtggtatactatgtctcagaacatttgcgtcca 5620
Db 301 GCATCCTGTCCCTGCTGCTTTCTGTGGGCTATACTATGTTCTCAGAACATTTGCTGCCA 360

QY 5621 ggaattgtgaacggggaagaggttggagcttgcaattttccgcaacttcactgtgagc 5680
Db 361 GGAATTTGTGAACGGGGAAGAGGTGTGGAGCTTGCATTTTTCACGCACTTCACTGTGGAGC 420

QY 5681 cggagctgcatttccataaaatcagagaatgccagtggtccctggttgaaggtaaaac 5740
Db 421 CGGAGTGTGCATTTTCTTAAATCAGAGAATGCCAGTGGTGGCTGCTGGTGAAGGTAAAGC 480

QY 5741 cagaggtgtgctatccagctccctacttggatgaataatgagaaacagacctggcct 5800
Db 481 CAGAGGCTGTGCTATCCAGCTCCCTTACTTGGATGAATATGAGAAACAGACCTGGCCT 540

QY 5801 gaagaggggcaaccccccttattatctgtgagcgggtatcggaagctccattgtgtcgt 5860
Db 541 GAAGAGGGGCAACCCCTTCATTTATCTCGTGAGGGGTATCGGAAGCTCCATTTGGTCTG 600

QY 5861 gcaacaacactcattatagaagatttctaggagcccaagagactaatcagatgtatt 5920
Db 601 GCAACAACACTCATTTATAGAAGAGATTGCTAGGAGCCCAAGAGACTTAATCAGATGTTATT 660
```

```
QY 5921 tggattcaactggcagttactg 5942
Db 661 GGGATTCACTCGCAGNACTG 682

RESULT 2
BG534574 756 bp mRNA EST 03-APR-2001
LOCUS 602553425F1 NIH_MGC_77 Homo sapiens cDNA clone IMAGE:4663182 5',
DEFINITION mRNA sequence.
ACCESSION BG534574
VERSION BG534574
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 756)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: CLONTECH Laboratories, Inc.
CDNA Library Preparation: CLONTECH Laboratories, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1465 row: a column: 07
High quality sequence stop: 751.
Location/Qualifiers
1..756
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4663182"
/clone_lib="NIH_MGC_77"
/lab_host="DH10B (T1 phage-resistant)"
/notes="Organ: lung; Vector: pDNR-LIB (Clontech); Site_1:
SfiI (ggccattatggcc); Site_2: SfiI (ggccattatggcc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CACGCCATTATGGCC-3' and 3' adaptor sequence:
5'-ATTCTAGAGCGAGCGCGCCGACATG-dt(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.9
kb (range 0.5-4.0 kb). 12/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: this is a NIH_MGC Library."
BASE COUNT 274 a 138 c 152 g 192 t
ORIGIN

Query Match 10.4%; Score 656.6; DB 11; Length 756;
Best Local Similarity 95.9%; Pred. No. 6.5e-143;
Matches 730; Conservative 0; Mismatches 19; Indels 12; Gaps 5;

QY 2836 ggtatgaactgcccagagcttttaacaagaccatctacaaaagaccagatttgatta 2895
Db 1 GGTATGAACCTTGGCAGGCTTTTAAACAGACCATTATCTACAAAGACCAGGATTTGATTA 60

QY 2896 aacaataataactaataagaagaatgcttcagggtcctcatctatattgtgggtgagc 2955
Db 61 AACAAATATAACTACTAATAGAGAAATGCTTCAGGTCTCATCTATATTGTTGGGTGAGC 120

QY 2956 gttatgtacctggagtggaataatgtaccaaaagagaggtcacatgagagaataatc 3015
Db 121 GTTATGTACCTGGAGTGGAAATGTGACCAAAGAGAGGTCCACATGAGAGAAATCAATTC 180

QY 3016 actgtcttgcattgaaccatgccacagctgcccattgcccataatttacctgagaatg 3075
Db 181 ACTTGTCTTCATTGAACCCCATGCACACAGTGCATTCGCAATTTGCCAAAATTTTACTCTGAGAATG 240
```



```

VERSION      BG862813.1  GI:14213351
KEYWORDS     EST.
SOURCE       Mus musculus
ORGANISM     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE    1 (bases 1 to 797)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: Lothar Hennighausen Ph.D., Priscilla Furth
             Ph.D.
             CDNA Library Preparation: Life Technologies, Inc.
             CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             http://image.llnl.gov
             Plate: LLAM10863 row: d column: 19
             High quality sequence stop: 690.

FEATURES     Location/Qualifiers
             1..797
             /organism="Mus musculus"
             /db_xref="taxon:10090"
             /clone="IMAGE:4934370"
             /clone_lib="NIH_CGAP_Mam4"
             /tissue_type="tumor, gross tissue"
             /lab_host="DH10B"
             /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: NotI;
             Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
             Average insert 2.5 kb. Library constructed by Life
             Technologies, catalog # 12018-016. Investigators providing
             samples: Lothar Hennighausen/Priscilla Furth, NIH
             Reference for transgenic model: Li et al., Cell Growth and
             Differentiation 7, 3-11 (1996). Note: This is a NCI_CGAP
             Library."

BASE COUNT   194 a 177 c 207 g 219 t
ORIGIN

Query Match          9.9%; Score 627.2; DB 11; Length 797;
Best Local Similarity 87.6%; Pred. No. 5e-136;
Matches 697; Conservative 0; Mismatches 98; Indels 1; Gaps 1;

QY 5171 tcttgctggtattgtgggtcctactgaagaatggcacccttatcttcgtgtgc 5230
DB 11 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2 tcccgctgggtaccgtggtgctctccctgaggaacggcatcaccccttacctccgtgtgc 61
QY 5231 tgcattgtttccactatttacttgggttaactccgctgaggaactgcataccaattc 5290
DB 11 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 62 tgcactgcttttccactatttacttgggtgagtagctccgctgaggaactgttgcctaattc 121
QY 5291 tgcagaagagagtagtcagtcactctgtagctatctatttacctacaataattgttct 5350
DB 11 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 122 tgcctgagagagatccagtcgactctgtgagctatctatttacctacaataattgttct 181
QY 5351 gctctccagggaattggtgactgaaggcccttgcctccagagtggtgtgcagatcc 5410
DB 182 gcttttccagggaattggtgactgaaggcccttgcctccagagtggtgtgcagatcc 241
QY 5411 tgccttactaaactgttgaagcaaaaaaacccgctggtcaggtacccttagaaaaagaaa 5470
DB 11 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 242 tgccttactaaactgttgaagcaaaaaaacccgctggtcaggtacccttagaaaaagaaa 301
QY 5471 tagttgatagactctcctgactatagctgctcctcctgaatcaagcttctcattcag 5530
DB 11 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 302 tagttgatagactctcctgactatagctgctcctcctgaatcaagcttctcattcag 361
QY 5531 gtgccacggtctgcagatgagcgaaagcatcctgctcctcctgcttcttctgtgggccc 5590
DB 11 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 362 gtgtccacgggtctgcagatgagcgaaagcatcctgctcctcctgcttcttctgtgggccc 421

```

---

```

QY 5591 tatactatgttctcagaacaatttgcctgcccaggaataattgtgaacggggaagaggttgagc 5650
DB 422 CATCCCTGTGTCTCAACAACATCTGTGCCAAGAAATAGTGAATGGGAAGAGGTGGAGC 481
QY 5651 ttgcatttttcacgcacttcacttcgtgagccgagctcgtattttctctctctctctctctct 5710
DB 482 GTGCGTTTTTCATCGCTTCATTTGGTGTGTGGAGTCTGCATTTTCTCTCTCTCTCTCTCTCT 541
QY 5711 atgccagtggtgctgctggttgaagtaaacgagagcgtgctgctctctctctctctctctct 5770
DB 542 ATGCAGGTGGTCTGTGTGAAGAAAGCCAGAGGCTGTGCCTACCCAGCCCTTACTT 601
QY 5771 gtagtaattgagagaacagaccctgctgagagggggaaggggggaacccctctctctctctct 5830
DB 602 GGATGAATATGGAGAAACAGACCCAGGCTTAAAGAGAGAGAAACCCACTTCTATTTATCTCG 661
QY 5831 tgagcgtgatcgaagctccatttggc-tggcaacaacactgcattatagaacagattg 5889
DB 662 GGAGCGGTATCGGAAGCTGCATTTGGTCTTTGGCAAAAGGACTGGATTATAGAAGAGATG 721
QY 5890 ctaggagccaagagactaatcagatgttatttgattcaactgagctgagcttctgtgagctc 5949
DB 722 GTCCGAGCCGGGAGACTAATCAGATGCTATTTGGACTTAAGTGGAGTACTCTCTGAGATC 781
QY 5950 caactctgctcctcaaga 5965
DB 782 CAGTTGTCTCAAGAGA 797

RESULT 5
BI086469 782 bp mRNA EST 20-JUN-2001
LOCUS 602849734F1 NIH_MGC_10 Homo sapiens cDNA clone IMAGE:4991193 5',
DEFINITION mRNA sequence.
ACCESSION BI086469
VERSION BI086469.1 GI:14504799
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 782)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: ATCC
             CDNA Library Preparation: Life Technologies, Inc.
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             http://image.llnl.gov
             Plate: LLAM11008 row: d column: 10
             High quality sequence stop: 715.

FEATURES     Location/Qualifiers
             1..782
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /clone="IMAGE:4991193"
             /clone_lib="NIH_MGC_10"
             /cell_line="MGC36"
             /lab_host="DH10B"
             /note="Organ: cervix; Vector: pCMV-SPORT6; Site_1: NotI;
             Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
             Average insert size 1.5 kb. Library prepared by Life
             Technologies."

BASE COUNT   184 a 194 c 163 g 241 t
ORIGIN

```



```

Db 85 GAGGCTGGAAAGTATTGGAGGGCTTTTGTCTCCATGTCAGGTTCACTTACATCAATA 26
Qy 6081 aaattttcttaattga 6097
Db 25 AAATATTTCTTAATGGA 9

RESULT 7
BE589438 606 bp mRNA EST 28-AUG-2000
LOCUS 195602 BARC 5BOV Bos taurus cDNA 5', mRNA sequence.
DEFINITION BE589438
ACCESSION BE589438
VERSION BE589438.1 GI:9842477
KEYWORDS EST.
SOURCE cow.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
Sonstegard,T.S., Capuco,A.V., Van Tassel,C.P., Ashwell,M.S. and
Wells,K.D.
Mapping of Expressed Sequence Tags from a normalized bovine mammary
gland cDNA library
Unpublished (2000)
Contact: Sonstegard TS
USDA, ARS, Beltsville Agricultural Research Center
Bldg. 200 Rm 2A, Beltsville, MD 20705, USA
Tel: 301 504 8416
Fax: 301 504 8414
Email: tads@psi.barc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.
PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCCTCCAGTCAGCG
Plate: 120 row: N column: 11
Seq primer: ATTTAGGTGACACTATAG.
FEATURES
source
1..606
Location/Qualifiers
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="BARC 5BOV"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: PCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled mRNA isolated from mammary
tissues at eight physiological, developmental, and disease
states."
BASE COUNT 144 a 146 c 113 g 203 t
ORIGIN
Query Match 8.1%; Score 512.2; DB 10; Length 606;
Best Local Similarity 90.4%; Pred. No. 4.2e-109;
Matches 547; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

Qy 4761 aaagcattaatgcagttgtcagtcacagagattacctgtcctcaggtctctgtacacag 4820
Db 1 AAAGCATTAATGCAGTTTGCCATTGCACAAAGGATTACCTGTTCTCAGGTCCTGTATACAG 60
Qy 4821 aaacatctggttcgtcttctatcagttgttcttctcctcaataaaatcagaagatacacca 4880
Db 61 AAACATCTGATTGCTGCTTCTATCAGTTGTCTCTCTAACTGCAATCAGAAGCTACACCT 120
Qy 4881 tgccctctctatagatctgtttcatgttttgggtgggtgctgtgttagcattcccatcc 4940
Db 121 TGCCCTCTATCCATAGATCTATTTCAAGTTTGGGGGGTACTGTGTAGCATTCCTCGTCC 180
Qy 4941 ttgtattgggatgacctgttgatctgcagcctcttctcagttagttcttctctataaccac 5000
Db 181 TTATACTGGGATGATGCTGTTCTGCTGTCAGCCTTACCAGTTAGCTTCTTCGTATAACCAC 240

```

```

Qy 5001 ctttatcttccatttgatcaccatggcacacatgcttcagatactactactacagtagac 5060
Db 241 CTTTATCTCTTCCATTGATCACCATGGCACACATGCTTCAGATACCTCTTACCATAGAC 300
Qy 5061 acagggctaccctctgctcaggttcaagaagacagtgtaagaggctcattccgcactctct 5120
Db 301 ACAGACCTCCCTTGGCAGAGATACAGGAGAGAGTGAAGAGGCTGTTCTGCATCTCTCT 360
Qy 5121 ttctttgcagaaattcttcaataatacaagtggtccattgggtgctgatatattcctggctgg 5180
Db 361 TTTCTTGGCAGAACTTCTCAGTATACAAAGTGGCTGATGGTGTGGTATTTCTTGGCTGG 420
Qy 5181 tatttggtggtcactgaagaatggtcatcaccccttatcttctgctgtgctgactgttt 5240
Db 421 TATTATGGTCTCTCTTGAAGAATGGCATCCCGTATCTTCGCTGTGCTGCTATGTTT 480
Qy 5241 ttccatatttacttggtgtaactccgcctgaggaactgcataccaattctgcagaagaa 5300
Db 481 TTCCACTATTACTTGGAGTAACCTCCACCTGGAAGAACTGTTTACCAATGCTGCAGAAAGA 540
Qy 5301 ggtacagtcactctgtagctatctatcttacctacaaattgttctcgtctcttccag 5360
Db 541 GAATACAATGCTACTCTGTAGCTATTTATCTTACCCACAATTTGTTCTCTGCTTTCCAG 600
Qy 5361 gaata 5365
Db 601 AAATA 605

RESULT 8
BG219270/c 624 bp mRNA EST 21-APR-2001
LOCUS RST39023 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
DEFINITION BG219270
ACCESSION BG219270
VERSION BG219270.1 GI:13745291
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 624)
AUTHORS Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R.,
Cain,S., Dahl,T., Thornton,M., Ramachandran,R., Whittington,J.,
Lerner,L., Krashoc,D., McElligott,K., Clark,S., Mays,R., Smith,E.,
Velooso,N., Hess,J., Cothren,K., Lo,K., Offenbacher,J., Danzig,J.,
and Ducar,M.
TITLE Creation of Genome-wide Protein Expression Libraries using Random
Activation of Gene Expression
JOURNAL Nat. Biotechnol. 19 (5), 440 (2001) In press
COMMENT Contact: Scott J. Cain
Athersys, Inc.
3201 Carnegie Ave, Cleveland, OH 44115, USA
Tel: 216 431 9900
Fax: 216 361 9596
Email: scai@atersys.com
High quality sequence stop: 624.
FEATURES
source
1..624
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Athersys RAGE Library"
/note="See 'Creation of Genome-wide Protein Expression'
Libraries using Random Activation of Gene Expression',
Nature Biotechnology, in press. Note that even though the
cell type indicated is HT1080, since a random activation
method was used, these sequence tags are not necessarily
expressed in HT1080 under normal circumstances."
BASE COUNT 129 a 151 c 136 g 206 t 2 others
ORIGIN

```

```
Query Match 7.6%; Score 479.8; DB 11; Length 624;
Best Local Similarity 98.2%; Pred. No. 1.6e-101;
Matches 484; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 3723 attactcatgataaagaaagcagaacgaaagaaagctgagctgtaggtacat 3782
Db 624 ATTACTCATGATAAAGAAAGCAGACGAAAGAAAGCTGNAGCTGNTAGTTACAT 565

QY 3783 gcgcagaagatcaggtcagatgtctgcttacagaaaaaacttcattgaaactcataa 3842
Db 564 CGCCAGAAAGATCATGGCTCAGATGCTGCTTACAGAAAAAATTCATTGAAACTCATAAA 505

QY 3843 ctcatgtatgacaatatacatcagaatgcctgggaaagaaagattccattatgagagaag 3902
Db 504 CTCATGTATGCAATATACATCAAGAAATGCTGGGAAGAAAGATTCCATTATGGAGGAAGAG 445

QY 3903 agcaccacagcagtcagtgactactctagaaattgctttgggtcctaaacggggtccatct 3962
Db 444 AGCACCCACAGTCAGTCAGTACTCTAGAAATGCTTTGGGTCTTAACGGGGTCCATCT 385

QY 3963 gttactgaaagagggtgctgacgtgcatccttggcgaagaaagacaggggtgaaata 4022
Db 384 GTTACTGAAAGAGGAGTGTGACGTGATCCTTTGCCAAGAAAGACAGGAGGTGAAATA 325

QY 4023 gaaaataatgccatgtattatcgctgtgtccagaaatctactgccttaaccagcac 4082
Db 324 GAAATAATGCCATPGGTATTATCGCCCTGTGTCCAGAAATCTACTGCCCTTAACCCAGCAC 265

QY 4083 aggggaaacccatagaaactctcaggagaagccctagaccacatttcagatccagac 4142
Db 264 AGGGAAACCCATAGAACTCTCAGGAGAACCCCTAGACCACCTTTTCATGGATCCAGAC 205

QY 4143 ttggcatatgaaactttacaggaagctgtggtcatgttaatgcgcgcagtgctgtggcag 4202
Db 204 TTGGCATATGGAACCTTATACAGGAAGCTGTGTCATGTAAATGCACGCGAGTGTGTGGCAG 145

QY 4203 aagtattttgaag 4215
Db 144 AACTCCTCCGGAG 132

RESULT 9
AII92195/c
LOCUS AII92195 481 bp mRNA EST 28-OCT-1998
DEFINITION qc92e08.x1 Soares_pregnant_uterus_NBHPU Homo sapiens cDNA clone
IMAGE:1721702 3' similar to TR:015057 O15057 KIAA0349 ;, mRNA
sequence.
ACCESSION AII92195
VERSION AII92195.1 GI:3743404
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 481)
NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 563 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 351.
Location/Qualifiers
1. .481
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1721702"
/clone_lib="Soares_pregnant_uterus_NBHPU"
/sex="female"

FEATURES
source
```

```
/dev_stage="adult"
/lab_host="DH10B"
/Note="Organ: uterus; Vector: pT73-Pac; Site.1: Not I;
Site.2: Eco RI; 1st strand cDNA was primed with a Not I -
oligo(dT) primer [5',
AACTGGAGAATTCGGCGCGCTTTTTTTTTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization. Library
constructed by M. Fatima Bonaldo."

BASE COUNT 125 a 96 c 86 g 174 t
ORIGIN

Query Match 7.6%; Score 477.8; DB 10; Length 481;
Best Local Similarity 99.6%; Pred. No. 4.6e-101;
Matches 479; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2778 cttcagattggtgcattcttaattggtatcccaataagttctgttactggtacttcagagg 2837
Db 481 CTTCAGATTGGTGCATCTTTAATGGATCCAAATAAGTCTTGTGTACTGGTACTTCAGAGG 422

QY 2838 tatgaacttgcgagggcttttaacaagaccatctctacaaaagaccggatttgattaaa 2897
Db 421 TATGAACCTTGGCGAGGCTTTTAAACAAGACCATATCTACAAAAGCCAGGATTTGATTAAA 362

QY 2898 caataataacactaataagaagaatgcttcagtcctcctatctatttgggtgagcgt 2957
Db 361 CAATAATAACACTAATAAGAAGAAATGCTTCAGGTCTCTCATATATTGTGGGTGAGCGT 302

QY 2958 tatgtacctggagtggaatgtgaccaagaagaggtcacaatgagagaatcatcac 3017
Db 301 TATGTACCTGGAGTGGGAAATGTGACCAAGAAGAGAGTCAACATGAGAGAATCATTCAC 242

QY 3018 ttgcttgcattgaacccatgcccacagtgccattgcccataaatttcactgagaatgaa 3077
Db 241 TTGCTTTGCATTGAACCCATGCCACAGTGCCATTGCCAAAAATTTACCTGAGAATGAA 182

QY 3078 aataatgaactggcttagaagaatgcataaacaagaagtgccacatttaagaaacccaggt 3137
Db 181 AATAATGAACCTGGCTTAGAGAATGCTAATAAAGTGGCCACATTTAAGAAACCCAGGT 122

QY 3138 gtatcaggccatggagctttatgaactaaagatgaatcactgaaagacttcaatgtac 3197
Db 121 GTATCAGGCCATGGAGTTTATGAACCTAAAGATGAATCACTGAAGAGACTTCAATATGTAC 62

QY 3198 tttatcattactccaaacccagcagatagcaaggctgaacatgacagaagaaagaga 3257
Db 61 TTTTATCATTTACTCCAAAACCCAGCATAGCAAGGCTGAACATATATGCAGAAAAGGAGA 2

QY 3258 a 3258
Db 1 A 1

RESULT 10
AW971391
LOCUS AW971391 583 bp mRNA EST 01-JUN-2000
DEFINITION EST383480 MAGE resequences, MAGL Homo sapiens cDNA, mRNA sequence.
ACCESSION AW971391
VERSION AW971391.1 GI:8161236
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 583)
Hegde,P., Qi,R., Abernathy,K., Dharap,S., Gaspard,R., Gay,C., Holt
,I.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and
Quackenbush,J.
Assessment of gene expression patterns in a model of colon tumor
metastasis using a 19,200 element cDNA microarray
```





```

RESULT 12
AW291190/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 522)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapb-r@mail.nih.gov
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. cDNA library Preparation: M.B. Soares Lab Clone Distribution:
NCI-CGAP clone distribution information can be found through the
I.M.A.G.E Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: M13 Forward
POLYA=Yes.

Location/Qualifiers
1..522
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2723895"
/clone_lib="NCI CGAP Sub4"
/lab_host="DH10B (Life Technologies)"
/notes="Vector: p773D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; The
NCI-CGAP_Sub4 library is a subtracted library derived from
the NCI-CGAP_Sub2 library which is a subtracted library
derived from the NCI-CGAP_Sub1 library, which is a
subtracted library derived from BI. BI constitutes a
mixture of 21 normalized or subtracted NCI-CGAP
libraries: NCI-CGAP_Co4, NCI-CGAP_Co16, NCI-CGAP_Pr28,
NCI-CGAP_Kid12, NCI-CGAP_Co16, NCI-CGAP_Kid5,
NCI-CGAP_Lym2, NCI-CGAP_Kid3, NCI-CGAP_Kid11,
NCI-CGAP_Le12, NCI-CGAP_Br2, NCI-CGAP_Co8, NCI-CGAP_CLL1,
NCI-CGAP_Lu24, NCI-CGAP_Brn23, NCI-CGAP_Lu5,
NCI-CGAP_Lu19, NCI-CGAP_GC4, NCI-CGAP_GC6,
NCI-CGAP_Brn25. These 21 libraries were pooled and a
single-stranded DNA preparation of the resulting mixture
was used as a tracer in a subtractive hybridization with
a driver whose composition is detailed below:
NCI-CGAP_Kid3 pool 1 : LLAM 3334-3337, 3682-3683,
3798-3803 (IMAGE Clonides 1322376-1323911,
1456008-1456775, 1500552-1502855) NCI-CGAP_Kid5 pool 1 :
LLAM 3338-3342, 3722-3725, 3776-3778 (IMAGE Clonides
1323912-1325831, 1471368-1472903, 1492104-1493255)
NCI-CGAP_Lu5 pool 1 : LLAM 3575-3582, 3851-3854 (IMAGE
Clonides 1414920-1417991, 1520904-1522439) NCI-CGAP_GC4
pool 1 : LLAM 3164-3167, 3716-3720, 3733-3735 (IMAGE
Clonides 1257096-1258631, 1469064-1470983, 1475592-1476743
) NCI-CGAP_Pr22 pool 1 : LLAM 2457-2459, 2758-2759.
3062-3068 (IMAGE Clonides 985608-986759, 1101182-1101959,
1217928-1220615) NCI-CGAP_Co10 pool 1 : LLAM 2644-2653,
2871-2872 (IMAGE Clonides 1057416-1061255, 1144584-1145351
) Subtraction was performed as previously described
[Bonaldo, Lennon & Soares (1996): Normalization and
Subtraction: Two Approaches To Facilitate Gene Discovery.
Genome Research 6, 791-806.]
TAG_LIB=NCI-CGAP_Co4
TAG_TISSUE=colon
TAG_SEQ=CTTCG"

BASE COUNT 137 a 125 c 100 g 160 t

```

## ORIGIN

```

Query Match 7.2%; Score 454; DB 10; Length 522;
Best Local Similarity 98.0%; Pred. No. 1.7e-95;
Matches 492; Conservative 0; Mismatches 5; Indels 5; Gaps 3;

QY 5596 tatgtttctcagaacatttgctgccaggaaattgtgaacgggggagaggttgagcttgcga 5655
|||||
Db 513 TATGTTCTCAGAACATTTGCTGCCAGGAAA-TGTGAACGGGGAAGAGTTGGAG---TTCA 458
|||||
QY 5656 tttttcacgcacttcactgtggagccggagctctgcattttcctaaaaatcagagaatgcc 5715
|||||
Db 457 TTTTTCACGCACACTTCACGTGTGGAGCCGAGTCTGCATTTTCTTAAAAATCAGAGAATGCC 398
|||||
QY 5716 gactggtccctggtgaaggttaaagccagagctgtcctatccagctccttacttggatg 5775
|||||
Db 397 GAGTGTCTCTGGTTGAAGGTAAAGCCAGAGGCTGTCTATCCAGCTCCTTACTTTGGATG 338
|||||
QY 5776 aatatggagaacacagaccctggcctgaaggggcaaccccttcatttctctgtgagc 5835
|||||
Db 337 AATATGGAGAAACAGACCCCTGGCTGAAGAGGGGCAACCCCTTCATTTATCTCGTGAGC 278
|||||
QY 5836 ggtatcggagctccatttggctgtgccaacacactgcattatagaagagattgctagga 5895
|||||
Db 277 GGTATCGGAAGCTCCATTTTGGTCTGGCAACAACACTGCATTTATAGAAGAGATTGCTAGGA 218
|||||
QY 5896 gccacagagactaatcagatgtatttggattccaactggcagttactgtgagctccaactc 5955
|||||
Db 217 GCCAAGAGACTAATCAGATGTTATTGGATTCAACTGGCAGTTACTGTGAGCTCCAACCTC 158
|||||
QY 5956 tgctcaagacacatcacaaatgcagcagtagtaaaagctgattcaaaaattatggaataac 6015
|||||
Db 157 TGCCTCAAGACATCAACAATGACGACAGTAGTAAGAGCTGATTCAAAATTTATGGAAAAAC 98
|||||
QY 6016 ttctgagggctgggaaagtattggagggtcttttcttctccatgtccaggttcacttaccat 6075
|||||
Db 97 TTTCTCAGGCTGGGAAGATTTTGGGA-GGTCTTTTGTCTCCATGTCCAGGTTCACTTACAT 39
|||||
QY 6076 caataaaaattttcttaatgga 6097
|||||
Db 38 CAATAAAATATTCTTAAATGGA 17
|||||

RESULT 13
BG382648 522 bp mRNA EST 12-MAR-2001
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
1 (bases 1 to 522)
AUTHORS
Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.
and Keese,J.W.
TITLE
Design and use of two pooled tissue normalized cDNA libraries for
EST discovery in swine
JOURNAL
Unpublished (2000)
COMMENT
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt.trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.
PCR primers
FORWARD: AGGAACACGCTATGACCAT

```

FEATURES		TITLE		Design and use of two pooled tissue normalized cDNA libraries for EST discovery in swine	
source		JOURNAL		Unpublished (2000)	
Seq primer: ATTTAGGTGACACTATAG.		COMMENT		Contact: Smith TPL USDA, ARS, US Meat Animal Research Center PO Box 166, Clay Center, NE 68933-0166, USA Tel: 402 762 4366 Fax: 402 762 4390 Email: smith@email.marc.usda.gov Single pass sequencing. Bases called and alt_trimmed with phred v0.980904.e. Vector identified by cross_match with the -minscore 18 and -minmatch 12 options. PCR Primers FORWARD: AGGAACACGCTATGACCAT BACKWARD: GTTTTCCAGTCACGACG Plate: 4 row: B column: 23 Seq primer: ATTTAGGTGACACTATAG.	
BASE COUNT		FEATURES		Location/Qualifiers	
ORIGIN		Source		1..520	
				/organism="Sus scrofa"	
				/db_xref="taxon:9823"	
				/clone_lib="MARC lPIG"	
				/tissue_type="pooled"	
				/lab_host="DH10B"	
				/note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI; Library made from pooled tissue from day 11, 13, 15, 20, and 30 embryos."	
122 a		121 a		126 c	
126 c		99 g		174 t	
Query Match		7.1%		Score 448; DB 11; Length 522;	
Best Local Similarity		91.4%		Pred. No. 2.6e-94;	
Matches 476;		Conservative 0;		Mismatches 45; Indels 0; Gaps 0;	
Qy 4919		tgctgtgttagcattcccatcctgtattgggatgaccctgttgatctgcagccctcttc		4978	
Db 1		TGCTGTGTTAGCTTCCCATCCTTCTATTGGGATGATGCTGTGTGATCTGCAGCCTTCCTC		60	
Qy 4979		agttagttcttcttaaccacctttatctctccattgtatgcacatgggcacacatgct		5038	
Db 61		AGTTAGTCTTCTCTATAACCACCTTTATCTCTCCATTGTGATCCACATGGCACATGTT		120	
Qy 5039		tcagatactacttacagtagacacagcagccctacccttgctcaggttcaagaagacagtg		5098	
Db 121		TCAGATACTTCTTACCATAGACACAGGCCCTTCCCTTTGCACAGACTCAAGAAGAGAGTGA		180	
Qy 5099		agaggtcattccgcacatctctcttcttgacagaaattctcaataatacaagtggtccat		5158	
Db 181		AGAGGCTCATTCGTCATCTCTTCTTGGCAGAGATTCTCAATATACAAATGGCTACGT		240	
Qy 5159		tggtgtgtgatattctctggtggttattgtggtgtcactgaagaatggcatcaccocctta		5218	
Db 241		TGGGTGTGGTATTCCTGGCTGGTATTTATGGGTCTCATTTGAAGAATGGCATCCCTTA		300	
Qy 5219		tcttcgtctgctcatgttttccactatttacttgggttaactccgcctgaggaact		5278	
Db 301		CCTTCGCTGTGCTGATTTTCCATTTTCCATTTATTTGGAATACTCCACCTGAAGATCT		360	
Qy 5279		gcataccaattctgcagaagagtagacagtgacactctgtagctatctatttaccctac		5338	
Db 361		GTTTACCATAATCTCGACAGAGCGAGTACAGTGCACACTCTGTAGCTATTATCTTTACCCAC		420	
Qy 5339		aaattgttctctcttccagggaattattgggatactgaaggcccttgctccagaggtg		5398	
Db 421		AAATTTGTTCTGCTTTTCCAGAAAATATTGGGATACTATAAGGCCCTCTCCAGAGGTG		480	
Qy 5399		gtgtgcagacctgccttactaaactgtttgaagcaaaaa		5438	
Db 481		GTGTGCAGACCCCTGCCTTAACTGTTTGAAGCAAAAA		520	
RESULT 14		RESULT 15			
BG382624		LOCUS		520 bp mRNA EST 12-MAR-2001	
DEFINITION		298590 MARC lPIG Sus scrofa cDNA 5', mRNA sequence.			
ACCESSION		BG382624			
VERSION		BG382624.1		GI:13307096	
KEYWORDS		EST.			
SOURCE		pig.			
ORGANISM		Sus scrofa			
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.			
AUTHORS		Fahrenkrug, S.C., Freking, B.A., Rohrer, G.A., Smith, T.P.L., Casas, E., Stone, R.T., Heaton, W.P., Grosse, W.M., Bennett, G.A., Laegreid, W.W. and Keele, J.W.			

BE650873/c  
 LOCUS BE650873 486 bp mRNA EST 06-SEP-2000  
 DEFINITION UI-M-BH3-ata-a-07-0-UI.r1 NIH\_BMAP\_M\_S4 Mus musculus cDNA clone  
 UI-M-BH3-ata-a-07-0-UI 5', mRNA sequence.  
 ACCESSION BE650873  
 VERSION BE650873.1 GI:9976697  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 486)  
 AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.  
 TITLE Normalization and subtraction: two approaches to facilitate gene  
 discovery  
 JOURNAL Genome Res. 6 (9), 791-806 (1996)  
 MEDLINE 9704477  
 COMMENT Contact: Chin, H  
 National Institute of Mental Health  
 6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD  
 20892-9643, USA  
 Tel: 301 443 1706  
 Fax: 301 443 9890  
 Email: mEst@mail.nih.gov  
 CDNA Library Preparation: M.B. Soares Lab Clone distribution:  
 Researchers may obtain BMAP cDNA clones from RESEARCH GENETICS. It  
 should be noted that Bento Soares is generating a small number of  
 additional specialized non-redundant arrays of BMAP cDNAs whose  
 availability will be considered under appropriate and limited  
 collaborative arrangements  
 Seq primer: M13 Reverse.  
 Location/Qualifiers  
 1..486  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UI-M-BH3-ata-a-07-0-UI"  
 /clone\_lib="NIH\_BMAP\_M\_S4"  
 /dev\_stage="27-32 days"  
 /lab\_host="DH10B (Life Technologies)"  
 /note="vector: pMT30-Pac (Pharmacia) with a modified  
 polylinker; Site\_1: Not I; Site\_2: Eco RI; The  
 NIH\_BMAP\_M\_S4 library is a subtracted library of a series,  
 ultimately derived from a mixture of individually tagged  
 normalized libraries from ten regions of the mouse brain  
 (cerebellum, brain stems, olfactory bulbs, hypothalamus,  
 cortex, amygdala, basal ganglia, pineal gland, striatum,  
 hippocampus) after a series of subtractions to reduce the  
 representation of cDNAs from which ESTs had already been  
 generated. The following serially subtracted libraries  
 were generated in this process: NIH\_BMAP\_M\_S4,  
 NIH\_BMAP\_M\_S3.3, NIH\_BMAP\_M\_S3.2, NIH\_BMAP\_M\_S3.1,  
 NIH\_BMAP\_M\_S2, NIH\_BMAP\_M\_S1. The subtracted library  
 (NIH\_BMAP\_M\_S4) was constructed as follows: PCR amplified  
 cDNA inserts from NIH\_BMAP\_M\_S3.3, NIH\_BMAP\_M\_S3.2, and  
 NIH\_BMAP\_M\_S3.1 clones from which 3' ESTs had been derived  
 was used as a driver in a hybridization with a pool of  
 the NIH\_BMAP\_M\_S3.3, NIH\_BMAP\_M\_S3.2, and NIH\_BMAP\_M\_S3.1  
 libraries in the form of single-stranded circles. The  
 remaining single-stranded circles (subtracted library)  
 was purified by hydroxyapatite column chromatography,  
 converted to double-stranded circles and electroporated  
 into DH10B bacteria (Life Technologies) to generate the  
 NIH\_BMAP\_M\_S4 library. This procedure has been previously  
 described (Bonaldo, Lennon and Soares, Genome Research  
 6:791-806, 1996)"

FEATURES  
source

QY 2079 caggggtatagccagcagacaaaattgggaagagtatatatgcagtaataatgatgacctaagaagtat 2138  
 |||||  
 Db 486 CAGGGGTATAGCCAGGACAACTGGGAAGAGTCTACGCAGTTATATGTGACCTAAAGTAT 427  
 |||||  
 QY 2139 atcctgatcagcaaacccacaataatgacagaaagattaaagattgcagttccttgaaggt 2198  
 |||||  
 Db 426 ATCCTGATTAGCAAGCCTGTCATATGACAGACAGATTAAGAGCGCAGTTCTCTGGAAGGT 367  
 |||||  
 QY 2199 ttctgatcttttttgaagattcttaccctgtatgcagggaatgggaagaatccgaagacag 2258  
 |||||  
 Db 366 TTCCGGTCTTTTCTGAAGATTCCTACCTGTATGTCAGGGAATGGAAGAATCAGAGACAA 307  
 |||||  
 QY 2259 gttgggcaacacattgaattggtatccctgattgggaggtgctccattgtctatcacagatcaa 2318  
 |||||  
 Db 306 GTTGGCAACACATTTGAAGTGGACCTGACTGGGAGGCTGCCATCGCTATACAGATGCAA 247  
 |||||  
 QY 2319 ttgaagaataattttactcatgttcccaagagtgtgctgtgtgatgaagaactcttactt 2378  
 |||||  
 Db 246 CTAAGAATAATTTTGGCTCATGTTCCTCAAGAGTGGTGTCTTGTGATGAAGATCTCTTACTG 187  
 |||||  
 QY 2379 gtgcttataaagaatgtcacaaagctgtgatgggtgcagttaccagtttcatatctagt 2438  
 |||||  
 Db 186 GTGGCTTATAAAGAAATGTCACAAAGCTGTATGAGGTGCAGTACAAATTTTCATGTCCAGT 127  
 |||||  
 QY 2439 agcaagacagtagtacaaatcgtgtggacatagtttggaaacaaagtcctacagagatctct 2498  
 |||||  
 Db 126 ACCAAGACAGTAGTGCATTTGTGGGTCTATAGTCTGGAACCAAAATCCTCAAAAGTGCT 67  
 |||||  
 QY 2499 gaggatctgtgaagcatacatctgccactctcttaggaccttctgtgtcttcatgacgt 2558  
 |||||  
 Db 66 GAGGACCTTGTGAAGCATACACCTGCCACTCTCTAGAACACTTGTGGTCTTTCATGTACGT 7  
 |||||  
 QY 2559 ttaagc 2564  
 |||||  
 Db 6 TTAAGC 1

Search completed: March 1, 2002, 14:46:36  
 Job time: 8041 sec

BASE COUNT 129 a 117 c 95 g 145 t  
 ORIGIN

Query Match 6.5%; Score 410.8; DB 10; Length 486;  
 Best Local Similarity 90.3%; Pred. No. 2.3e-85;  
 Matches 439; Conservative 0; Mismatches 47; Indels 0; Gaps 0;